PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/705, 14/725, A61K 38/17, C07K 16/28, G01N 33/68, C12Q 1/68

A2

(11) International Publication Number: WO 99/41375

(43) International Publication Date: 19 August 1999 (19.08.99)

(21) International Application Number: PCT/US99/02572

(22) International Filing Date: 5 February 1999 (05.02.99)

(30) Priority Data:
09/022,939 12 February 1998 (12.02.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/022,939 (CIP)
Filed on 12 February 1998 (12.02.98)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Montain View, CA 94043 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). AU-YOUNG, Janice [US/US]; 1419 Kains Avenue, Berkeley, CA 94702 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). SHAH, Purvi [IN/US]; 1608 Queen Charlotte Drive #5, Sunnyvale, CA 94087 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). BAUGHN, Mariah [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
*Without international search report and to be republished upon receipt of that report.*

(54) Title: HUMAN RECEPTOR PROTEINS

(57) Abstract

The invention provides human receptor proteins (HURP) and polynucleotides which identify and encode HURP. The invention also

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# HUMAN RECEPTOR PROTEINS

## TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human receptor proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cancer, autoimmune/inflammatory disorders, reproductive disorders, gastrointestinal disorders, and developmental disorders.

## BACKGROUND OF THE INVENTION

The term receptor describes proteins that specifically recognize other molecules. The category is broad and includes proteins with a variety of functions. The bulk of the proteins termed receptors are cell surface proteins which bind extracellular ligands, leading to cellular responses including growth, differentiation, endocytosis, and immune response. Other proteins termed receptors facilitate the specific transport of proteins across the endoplasmic reticulum membrane and localize enzymes to a particular location in the cell.

Cell surface receptors are typically integral membrane proteins of the plasma membrane. These receptors recognize hormones such as catecholamines, e.g., epinephrine, norepinephrine, and histamine; peptide hormones, e.g., glucagon, insulin, gastrin, secretin, cholecystokinin, adrenocorticotropic hormone, follicle stimulating hormone, luteinizing hormone, thyroid stimulating hormone, parathyroid hormone, and vasopressin; growth and differentiation factors, e.g., epidermal growth factor, fibroblast growth factor, transforming growth factor, insulin-like growth factor, platelet-derived growth factor, nerve growth factor, colony-stimulating factors, and erythropoietin; cytokines, e.g., chemokines, interleukins, interferons, and tumor necrosis factor; small peptide factors such as bombesin, oxytocin, endothelin, angiotensin II, vasoactive intestinal peptide, and bradykinin; neurotransmitters such as neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, e.g., enkephalins, endorphins and dynorphins; galanin, somatostatin, and tachykinins; and circulatory system-borne signaling molecules, e.g., angiotensin, complement, calcitonin, endothelins, and formyl-methionyl peptides. Cell surface receptors on immune system cells recognize antigens, antibodies, and major

histocompatibility complex (MHC)-bound peptide. Other cell surface receptors bind ligands to be internalized by the cell. This receptor-mediated endocytosis functions in the uptake of low density lipoproteins (LDL), transferrin, glucose- or mannose-terminal glycoproteins, galactose-terminal glycoproteins, immunoglobulins, phosphovitellogenins, fibrin, proteinase-inhibitor complexes, plasminogen activators, and thrombospondin. (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York, NY, p. 723; and Mikhailenko, I. et al. (1997) J. Biol. Chem. 272:6784-6791.)

Many cell surface receptors have seven transmembrane regions, with an extracellular N-terminus that binds ligand and a cytoplasmic C-terminus that interacts with G proteins. (Strosberg, A.D. (1991) Eur. J. Biochem. 196:1-10.) Cysteine-rich domains are found in two families of cell surface receptors, the LDL receptor family and the tumor necrosis factor receptor/nerve growth factor (TNFR/NGFR) receptor family. Seven successive cysteine-rich repeats of about forty amino acids in the N-terminal extracellular region of the LDL receptor form the binding site for LDL and calcium; similar repeats have been found in vertebrate very low density lipoprotein receptor, vertebrate low-density lipoprotein receptor-related protein 1 (LRP1) (also known as $\alpha_2$-macroglobulin receptor), and vertebrate low-density lipoprotein receptor-related protein 2 (also known as gp330 or megalin) (ExPASy PROSITE document PDOC00929; and Bairoch, A. et al. (1997) Nucl. Acids. Res. 25:217-221.) The structure of the repeat is a β-hairpin followed by a series of β-turns; there are six disulfide-bonded cysteines within each repeat.

In the extracellular regions of TNFR and NGFR there are cysteine-rich domains composed of three to four modules of about forty residues each. Each module contains six conserved disulfide-bonded cysteines. These repeats are found in many receptors including

TNFR types I and II, lymphotoxin alpha/beta receptor, low-affinity NGFR, and cytokine receptors CD40, CD27, and CD30. (ExPASy PROSITE document PDOC00561; and Bairoch, et al. supra.)

The frizzled cell surface receptor, originally identified in Drosophila melanogaster, is important for proper bristle and hair polarity on the wing, leg, thorax,

N-terminal signal sequence and seven putative transmembrane regions. The N-terminus is cysteine-rich and is probably located extracellularly while the C-terminus is probably cytosolic. Multiple frizzled gene homologs have been found in rat, mouse, and human. The frizzled receptors are not homologous to other seven-transmembrane-region receptors. The ligands that bind the frizzled receptors are still unknown.

Calcitonin gene-related peptide is a neuropeptide with multiple effects including vasodilation, neuromodulation, and inhibition of smooth muscle, including uterine, contraction. (Luebke, A.E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:3455-3460; and Naghashpour, M. et al. (1997) Endocrinology 138:4207-4214.) The calcitonin gene-related peptide receptor component protein (CGRP-RCP) was cloned from guinea pig cochlea (hearing organ) as a protein which confers CGRP responsiveness to Xenopus oocytes. (Luebke, et al. supra.) CGRP-RCP is co-localized with CGRP-containing efferent nerve terminals in guinea pig cochlea. A highly related mouse CGRP-RCP was identified whose levels drop during parturition, suggesting that this protein has a role in regulating uterine contractions during pregnancy. (Naghashpour, et al. supra.) Because of CGRP-RCP's hydrophilicity, researchers speculate that it is not itself the receptor, but rather associates with other proteins to form the receptor. Two seven-transmembrane-domain proteins have been identified which act as CGRP receptors. (Naghashpour, et al., supra.)

T cells play a central role in the immune system as effectors and regulators, coupling antigen recognition and the transmission of signals that induce cell death in infected cells and stimulate other immune cells. T cells recognize a wide range of different antigens, but a particular clonal line of T cells can only recognize a single antigen. T cells recognize antigens only when they are presented to the T cell receptor (TCR) as peptides in a complex with major histocompatibility molecules (MHC) on the surface of antigen presenting cells. The TCR on most T cells consists of immunoglobulin-like integral membrane glycoproteins containing two polypeptide subunits, $\alpha$ and $\beta$, of similar molecular weight. The TCR $\beta$ subunit has an extracellular domain containing both variable and constant regions, a transmembrane domain that traverses the membrane once, and a short intracellular domain. (Saito, H. et al. (1984) Nature 309:757-762.) The genes for the TCR subunits are constructed through somatic rearrangement of different gene segments. Interaction of antigen in the proper MHC

context with the TCR initiates signalling cascades that induce the proliferation, maturation, and function of cellular components of the immune system. (Weiss, A. (1991) Annu. Rev. Genet. 25: 487-510.)

The TCR antigen repertoire is established by developmentally regulated TCR gene rearrangements. Immature T cells undergo a selection and differentiation process based on antigen binding prior to leaving the thymus. T cells that bind self-antigens while still in the cortex of the thymus are eliminated by apoptosis, establishing immunological tolerance. Failure to eliminate self reactive populations of cells has been shown to result in autoimmune disease. (Olive, C. (1995) Immunol. Cell Biol. 73: 297-307.) Mice with mutant TCR β or α subunits develop an inflammatory bowel disease similar to human ulcerative colitis. (Mombaerts, P, et al. (1993) Cell 75:275-282.) Rearrangements in TCR genes and alterations in TCR expression have been noted in lymphomas, leukemias, autoimmune disorders, and immunodeficiency disorders. (Aisenberg, A.C. et al. (1985) N. Engl. J. Med. 313:529-533; Weiss, supra; and Olive, supra.)

A family of four-transmembrane span proteins unrelated to the tetraspanins exists that function as receptors in the immune system. This family includes the high affinity immunoglobulin E (IgE) receptor β subunit, the CD20 protein, and the HTm4 protein. Upon exposure to allergens the B cells of responsive individuals secrete IgE molecules specific to the allergen. IgE molecules bind to the high affinity IgE receptor (FcRI) present on mast cells and basophils. IgE binding activates the release of a variety of vasoactive mediators which promote allergic and inflammatory responses. (Beaven, M.A. and Baumgartner, R.A. (1996) Curr. Opin. Immunol. 8:766-772.)

The FcRI is a tetrameric complex composed of an α-chain, a β-chain, and a dimer of identical disulfide-linked γ-chains. IgE molecule binds through their Fc region to the FcRI α-chain with a 1:1 stoichiometry. The β- and γ-chains of FcRI are required for signal transduction upon activation of the receptor complex by IgE binding. The β-chain is associated with the tyrosine kinase Lyn, and the γ-chain is associated with the tyrosine kinase Syk. (Beaven and Baumgartner, supra.). Both the amino and carboxy termini of FCRIβ are probably located in the cytoplasm. (Kinet, J. P. et al. (1988) Proc. Natl. Acad. Sci. USA 85: 6483-6487.)

CD20, a related four-transmembrane-region protein, is expressed on the B cell surface where it initiates intracellular signals and modifies B cell growth and

differentiation. (Tedder, T.F. et al. (1988) Proc. Natl. Acad. Sci. USA 85:208-212.) CD20 is associated with the tyrosine kinases Lyn, Fyn, and Lck and with p75/p80 protein. Researchers hypothesize that CD20 may be the signal-transducing component of a receptor complex. (Deans, J.P. et al. (1995) J. Biol. Chem. 270:22632-22638.) Human HTm4 is expressed specifically in hematopoietic cells. (Adra, C. N. et al. (1994) Proc. Natl. Acad. Sci. USA 91: 10178-10182.) The human HTm4, CD20, and FcRI genes map to chromosome 11q12-13.1 and the murine homologue for CD20, Ly-44, and the murine FcRI β subunit map to the same region of murine chromosome 19. (Adra et al., supra.) Genetic linkage studies indicate that genes in this region show a close linkage with atopy, a sustained IgE response to allergens, and bronchial hyperreactivity. (Sandford, A. J. et al. (1993) Lancet 341: 332-334.)

Secreted polypeptides; integral membrane glycoproteins of the plasma, nuclear, rough endoplasmic reticulum (ER), Golgi, lysosome, and endosome membranes; and enzymes of the lysosome, rough ER, and Golgi complex are synthesized with N-terminal signal sequences on cytosolic ribosomes. The signal recognition particle (SRP), a ribonucleoprotein, binds the signal sequence after it emerges from the ribosome. The nascent chain/ribosome/SRP complex binds the SRP receptor on the cytosolic side of the ER membrane. The SRP dissociates from the signal sequence, allowing the nascent chain to begin translocating across the ER membrane. The SRP then dissociates from the SRP receptor, and both go on to catalyze new rounds of nascent chain targeting. The SRP receptor is composed of two subunits, SRα, a peripheral membrane protein, and SRβ, an integral membrane protein with one transmembrane region. (Miller, J.D. et al. (1995) J. Cell Biol. 128:273-282.) The dissociation of SRP from the SRP receptor is driven by GTP hydrolysis; SRα, SRβ, and the SRP54 protein are all GTPases.

Abnormal hormonal secretion is linked to disorders including diabetes insipidus (vasopressin), hyper- and hypoglycemia (insulin, glucagon), Grave's disease and goiter (thyroid hormone), and Cushing's and Addison's diseases (adrenocorticotropic hormone; ACTH). Cancer cells secrete excessive amounts of hormones or other biologically active peptides. Disorders related to excessive secretion of biologically active peptides by tumor cells include fasting hypoglycemia due to increased insulin secretion from insulinoma-islet cell tumors; hypertension due to increased epinephrine and norepinephrine secreted from pheochromocytomas of the adrenal medulla and sympathetic paraganglia; and carcinoid

syndrome, which includes abdominal cramps, diarrhea, and valvular heart disease, caused by excessive amounts of vasoactive substances (serotonin, bradykinin, histamine, prostaglandins, and polypeptide hormones) secreted from intestinal tumors. Tumors may exhibit ectopic synthesis and secretion of biologically active peptides, including ACTH and vasopressin in lung and pancreatic cancers; parathyroid hormone in lung and bladder cancers; calcitonin in lung and breast cancers; and thyroid-stimulating hormone in medullary thyroid carcinoma.

Ligatin is a filamentous peripheral membrane protein and lectin that acts as a trafficking receptor for phosphoglycoproteins (PGPs) including N-acetyl-β-glucosaminidase, acetylcholinesterase, and β-hexosaminidase. Ligatin localizes PGPs within endosomes and at the cell periphery where PGPs participate in specific metabolic processes and in intracellular adhesion. (Jakoi, E.R. et al. (1976) J. Cell Biol. 70:97-111; Jakoi, E.R. and Marchase, R.B. (1979) J. Cell Biol. 80:643-650; Gaston, S.M. et al. (1982) J. Cell Biochem. 18:447-459; and Malnar-Dragojevic, et al. (1997) Genomics 40:192-193.) Bound palmitic acid may be involved in the ligatin's membrane attachment. (Jakoi, E.R. et al. (1987) J. Biol. Chem. 262:1300-1304.) Ligatin was originally found in ordered fibrillar arrays on the surface of suckling rat ileum and has subsequently been found in mammalian and avian tissues during embryonic development and early differentiated states. Researchers hypothesize that, in suckling rat ileum, ligatin localizes n-acetyl-β-glucsoaminidase for the extracellular digestion of carbohydrate moieties in maternal milk. The ligatin gene has been cloned from mouse. (Malnar-Dragojevic et al., supra.)

The discovery of new human receptor proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of cancer, autoimmune/inflammatory disorders, reproductive disorders, gastrointestinal disorders, and developmental disorders.

## SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, human receptor proteins, referred to collectively as "HURP" and individually as "HURP-1," "HURP-2," "HURP-3," "HURP-4," "HURP-5," "HURP-6," "HURP-7," and "HURP-8." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID

NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8.

The invention further provides a substantially purified variant having at least 90% amino acid identity to the amino acid sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or to a fragment of any of these sequences. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID

fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, a fragment of SEQ ID NO 9, a fragment of SEQ ID NO:10, a fragment of SEQ ID NO:11, a fragment of SEQ ID NO:12, a fragment of SEQ ID NO:13, a fragment of SEQ ID NO:14, a fragment of SEQ ID NO:15, and a fragment of SEQ ID NO:16. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, a fragment of SEQ ID NO 9, a fragment of SEQ ID NO:10, a fragment of SEQ ID NO:11, a fragment of SEQ ID NO:12, a fragment of SEQ ID NO:13, a fragment of SEQ ID NO:14, a fragment of SEQ ID NO:15, and a fragment of SEQ ID NO:16, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, a fragment of SEQ ID NO 9, a fragment of SEQ ID NO:10, a fragment of SEQ ID NO:11, a fragment of SEQ ID NO:12, a fragment of SEQ ID NO:13, a fragment of SEQ ID NO:14, a fragment of SEQ ID NO:15, and a fragment of SEQ ID NO:16.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID

amin acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8 in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8, as well as a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8.

The invention also provides a method for treating or preventing an

autoimmune/inflammatory disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8.

The invention also provides a method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:7.

The invention also provides a method for treating or preventing a gastrointestinal disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:7.

The invention also provides a method for treating or preventing a gastrointestinal disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:8 and a fragment of SEQ ID NO:8.

The invention also provides a method for treating or preventing a developmental disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, a fragment of SEQ ID NO:1, and a fragment of SEQ ID NO:2.

The invention also provides a method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of

SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment f SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8 in a biological sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide sequence encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8 to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample. In one aspect, the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those

described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

**DEFINITIONS**

"HURP," as used herein, refers to the amino acid sequences of substantially purified HURP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to HURP, increases or prolongs the duration of the effect of HURP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of HURP.

An "allele" or an "allelic sequence," as these terms are used herein, is an alternative form of the gene encoding HURP. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HURP, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same HURP or a polypeptide with at least one functional characteristic of HURP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HURP, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HURP. The encoded protein may also be "altered," and may contain deletions,

insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent HURP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of HURP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments", "immunogenic fragments", or "antigenic fragments" refer to fragments of HURP which are preferably about 5 to about 15 amino acids in length and which retain some biological activity or immunological activity of HURP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to HURP, decreases the amount or the duration of the effect of the biological or immunological activity of HURP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of HURP.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fa, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind HURP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a

mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to a specific nucleic acid sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic HURP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the

strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding HURP or fragments of HURP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW™ Fragment Assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence.

As used herein, the term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding HURP, by northern analysis is indicative of the presence of nucleic acids encoding HURP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding HURP.

A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of HURP, of a polynucleotide sequence encoding HURP, or of a polynucleotide sequence complementary to a polynucleotide sequence encoding HURP. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl,

acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "homology," as used herein, refers to a degree of complementarity. There may be partial homology or complete homology. The word "identity" may substitute for the word "homology." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% homology or identity). In the absence of non-specific binding, the substantially homologous sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN™ program (LASERGENE™ software package, DNASTAR, Inc., Madison WI). The MEGALIGN™ program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in

sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no homology between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be calculated by the clustal method, or by other methods known in the art, such as the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization," as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term "hybridization complex" as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., $C_0t$ or $R_0t$ analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition," as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray," as used herein, refers to an array of distinct polynucleotides or oligonucleotides arranged on a substrate, such as paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The term "modulate," as it appears herein, refers to a change in the activity of HURP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HURP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to an oligonucleotide, nucleotide, polynucleotide, or any fragment thereof, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which are greater than about 60 nucleotides in length, and most preferably are at least about 100 nucleotides, at least about 1000 nucleotides, or at least about 10,000 nucleotides in length.

The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the encoded polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially equivalent to the terms "amplimers," "primers," and "oligomers" as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary

single stranded DNA and RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63.)

The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding HURP, or fragments thereof, or HURP itself may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein recognized by the binding molecule (i.e., the antigenic determinant or epitope). For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

As used herein, the term "stringent conditions" refers to conditions which permit hybridization between polynucleotide sequences and the claimed polynucleotide sequences. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37°C to 42°C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30°C to 35°C. In particular, hybridization could occur under high stringency conditions at 42°C in 50% formamide, 5X SSPE, 0.3% SDS, and 200 $\mu$g/ml sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency c nditions as described above, but in 35% formamide at a reduced temperature of 35°C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, and refers to cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of HURP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE™ software.

**THE INVENTION**

The invention is based on the discovery of new human receptor proteins (HURP), the polynucleotides encoding HURP, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, autoimmune/inflammatory disorders, reproduct

disorders, gastrointestinal disorders, and developmental disorders. Table 1 shows the sequence identification numbers, Incyte Clone identification number, and cDNA library for each of the human receptor proteins disclosed herein.

Table 1

| PROTEIN | NUCLEOTIDE | CLONE ID | LIBRARY |
|---|---|---|---|
| SEQ ID NO:1 | SEQ ID NO:9 | 29167 | SPLNFET01 |
| SEQ ID NO:2 | SEQ ID NO:10 | 150629 | FIBRANT01 |
| SEQ ID NO:3 | SEQ ID NO:11 | 611082 | COLNNOT01 |
| SEQ ID NO:4 | SEQ ID NO:12 | 1223275 | COLNTUT02 |
| SEQ ID NO:5 | SEQ ID NO:13 | 1255202 | MENITUT03 |
| SEQ ID NO:6 | SEQ ID NO:14 | 1261646 | SYNORAT05 |
| SEQ ID NO:7 | SEQ ID NO:15 | 2083528 | UTRSNOT08 |
| SEQ ID NO:8 | SEQ ID NO:16 | 1451415 | PENITUT01 |

Nucleic acids encoding the HURP-1 of the present invention were first identified in Incyte Clone 29167 from the fetal spleen tissue cDNA library (SPLNFET01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:9, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 29167 (SPLNFET01), 36294 (HUVENOB01), 1458963 (COLNFET02), 2263065 (UTRSNOT02), and 2693656 (LUNGNOT23).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1. HURP-1 is 584 amino acids in length and has one potential N-glycosylation site at N181; four potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at S313, S318, S404, and T505; seven potential casein kinase II phosphorylation sites at S88, S152, T171, S237, T259, T366, and S404; and four potential protein kinase C phosphorylation sites at S17, S246, S313, and S500. HURP-1 has chemical and structural homology with mouse ligatin (GI 1377880). In particular, HURP-1 and mouse ligatin share 78% identity. The most useful fragment of SEQ ID NO: 9 is from about nucleotide 725 to about nucleotide 739. Northern analysis shows the expression of this sequence in various libraries, at least 64% of which are immortalized or cancerous, at least 18% of which are from inflamed tissue, and at least 11% of which are from fetal/proliferating cells. Of particular note is the expression of HURP-1 in libraries

Incyte Clone 150629 from the untreated ataxia telangiectasia fibroblast cell line cDNA library (FIBRANT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:10, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 150629 (FIBRANT01), 639489 (BRSTNOT03), 878146 (LUNGAST01), and 995140 (KIDNTUT01), and shotgun sequences SAAB00536 and SAIA03775.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2. HURP-2 is 706 amino acids in length and has two potential N-glycosylation sites at N38 and N352; two potential cAMP- and cGMP-dependent protein kinase sites at T150 and S316; eight potential casein kinase II phosphorylation sites at T84, T499, T576, T585, T591, S596, S641, and S673; eight potential protein kinase C phosphorylation sites at S496, S519, S548, S596, T607, S645, S653, and S679; and one potential tyrosine kinase phosphorylation site at Y103. HURP-2 has a potential signal sequence from M1 to about G18. HURP-2 has chemical and structural homology with mouse frizzled 6 putative transmembrane receptor (GI 1151256). In particular, HURP-2 and mouse frizzled 6 putative transmembrane receptor share 83% identity. By homology to mouse frizzled 6 putative transmembrane receptor, the extracellular domain of HURP-2 is from amino acid residue 1 to about amino acid residue 199, the transmembrane domain is from about amino acid residue 200 to about amino acid residue 496, and the intracellular domain from about amino acid residue 497 to amino acid residue 706. The most useful fragment of SEQ ID NO: 10 is from about nucleotide 178 to about nucleotide 198. Northern analysis shows the expression of this sequence in various libraries, at least 42% of which are immortalized or cancerous, at least 32% of which are from inflamed tissue, and at least 16% of which are from fetal/proliferating cells. Of particular note is the expression of HURP-2 in libraries made from reproductive tissues (32%), and cardiovascular tissues (26%).

Nucleic acids encoding the HURP-3 of the present invention were first identified in Incyte Clone 611082 from the [illegible] DNA library (COLNNOT16)

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:3. HURP-3 is 267 amino acids in length and has four potential N-glycosylation sites at N13, N228, N232, and N260; three potential casein kinase II phosphorylation sites at S8, S115, and S210; and three potential protein kinase C phosphorylation sites at S3, S115, and S198. HURP-3 has chemical and structural homology with rat IgE receptor β-subunit protein (GI 204117) and human CD20 protein (GI 179308). In particular, HURP-3 shares 23% identity with rat IgE receptor β-subunit protein and 19% identity with human CD20 protein. By homology to rat IgE receptor β-subunit protein and human CD20 protein, the N-terminal intracellular domain of HURP-3 is from amino acid residue 1 to about amino acid residue 89, the transmembrane domain is from about amino acid residue 90 to about amino acid residue 224, and the C-terminal intracellular domain is from about amino acid residue 225 to amino acid residue 267. The most useful fragment of SEQ ID NO: 11 is from about nucleotide 278 to about nucleotide 301. Northern analysis shows the expression of this sequence in various libraries, at least 57% of which are immortalized or cancerous and at least 43% of which are from inflamed tissue. Of particular note is the expression of HURP-3 in libraries made from small intestine and colon tissues (93%).

Nucleic acids encoding the HURP-4 of the present invention were first identified in Incyte Clone 1223275 from the colon tumor tissue cDNA library (COLNTUT02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:12, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 34843 (THP1NOB01), 1223275 (COLNTUT02), 2445731 (THP1NOT03), 1724713(PROSNOT14), and 1773030 (MENTUNON3).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:4. HURP-4 is 248 amino acids in length and has three potential N-glycosylation sites at N8, N20, and N74; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at T110; two potential casein kinase II phosphorylation sites at S22 and T193; five potential protein kinase C phosphorylation sites at S36, T105, S126, T177, and S233; and a potential tyrosine kinase

shares 22% identity with rat IgE receptor β-subunit protein and 19% homology with human CD20 protein. By homology to rat IgE receptor β-subunit protein and human CD20 protein, the N-terminal intracellular domain of HURP-4 is from amino acid residue 1 to about amino acid residue 46, the transmembrane domain from about amino acid residue 47 to about amino acid residue 204, and the C-terminal intracellular domain from about amino acid residue 205 to amino acid residue 248. The most useful fragment of SEQ ID NO: 12 is from about nucleotide 215 to about nucleotide 232. Northern analysis shows the expression of this sequence in various libraries, at least 45% of which are immortalized or cancerous and at least 53% of which are from inflamed tissue. Of particular note is the expression of HURP-4 in libraries made from hematopoietic/immune tissues (21%) and gastrointestinal tissues (20%).

Nucleic acids encoding the HURP-5 of the present invention were first identified in Incyte Clone 1255202 from the brain meningioma tissue cDNA library (MENITUT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:13, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 348357 (LVENNOT01) and 1255202 (MENITUT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:5. HURP-5 is 130 amino acids in length and has one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S30; three potential casein kinase II phosphorylation sites at S12, T79, and T93; and four potential protein kinase C phosphorylation sites at S30, T49, S75, and T128. HURP-5 has chemical and structural homology with mouse CGRP-receptor component protein (GI 2599066). In particular, HURP-5 and mouse CGRP-receptor component protein share 85% identity. The most useful fragment of SEQ ID NO: 13 is from about nucleotide 274 to about nucleotide 297. Northern analysis shows the expression of this sequence in various libraries, at least 67% of which are immortalized or cancerous and at least 17% of which are from inflamed tissue.

Nucleic acids encoding the HURP-6 of the present invention were first identified in Incyte Clone 1261646 from the knee synovial tissue cDNA library (SYNORAT05) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID

(PROSTUT08), 1704186 (DUODNOT02), and 3394566 (LUNGNOT28).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:6. HURP-6 is 271 amino acids in length and has one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S63; three potential casein kinase II phosphorylation sites at S95, T114, and S213; ten potential protein kinase C phosphorylation sites at S6, T25, S59, S63, S75, S123, S135, S189, T203, and S247; one potential tyrosine kinase phosphorylation site at Y160; and one potential ATP/GTP-binding site motif A (P-loop) from G71LCDSGKT. HURP-6 has chemical and structural homology with mouse signal recognition particle receptor β subunit (GI 600886). In particular, HURP-6 and mouse signal recognition particle receptor β subunit share 90% identity. By homology to mouse signal recognition particle receptor β subunit, the ER-lumen-facing domain of HURP-6 is from amino acid residue 1 to about amino acid residue 35, the transmembrane domain from about amino acid residue 36 to about amino acid residue 54, and the cytoplasmic domain from about amino acid residue 55 to amino acid residue 271. The most useful fragment of SEQ ID NO: 14 is from about nucleotide 128 to about nucleotide 148. Northern analysis shows the expression of this sequence in various libraries, at least 38% of which are immortalized or cancerous and at least 38% of which are from inflamed tissue. Of particular note is the expression of HURP-6 in libraries made from reproductive tissues (23%) and gastrointestinal tissues (21%).

Nucleic acids encoding the HURP-7 of the present invention were first identified in Incyte Clone 2083528 from the uterine tissue cDNA library (UTRSNOT08) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:15, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 757728 (BRAITUT02), 827431 (PROSNOT06), 917601 (BRSTNOT04), 1314605 (BLADTUT02), 1334893 (COLNNOT13), 1346609

potential tyrosine kinase phosphorylation site at Y413. HURP-7 has a potential signal sequence from M1 to about A16; two potential TNFR/NGFR cysteine-rich regions from C140 to C174 and from C381 to C418; and three potential LDL receptor class A domains from L138 to S176, from R313 to G355, and from R397 to Y435. The most useful fragment of SEQ ID NO: 15 is from about nucleotide 280 to about nucleotide 297. Northern analysis shows the expression of this sequence in various libraries, at least 51% of which are immortalized or cancerous and at least 37% of which are from inflamed tissue. Of particular note is the expression of HURP-7 in libraries made from reproductive (30%), gastrointestinal (16%), and cardiovascular (15%) tissues.

Nucleic acids encoding the HURP-8 of the present invention were first identified in Incyte Clone 1451415 from the penile tumor tissue cDNA library (PENITUT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1451415 (PENITUT01), 2918942 (THYMFET03), and 3603481 (DRGTNOT01), and shotgun sequence SBJA02297.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:16. HURP-8 is 314 amino acids in length and has one potential N-glycosylation site at N205; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S312; four potential casein kinase II phosphorylation sites at T101, T133, S152, and S239; and five potential protein kinase C phosphorylation sites at S26, S81, T95, T159, and S212. HURP-8 has chemical and structural homology with human T-cell receptor β (GI 1100182). In particular, HURP-8 and human T-cell receptor β share 87% identity. By homology to human T-cell receptor β, the extracellular domain of HURP-8 is from amino acid residue 1 to about amino acid residue 285, the transmembrane domain from about amino acid residue 286 to about amino acid residue 307, and the intracellular domain from about amino acid residue 308 to amino acid residue 314. The most useful fragment of SEQ ID NO:16 is from about nucleotide 106 to about nucleotide 129. Northern analysis shows the expression of this sequence in various libraries, at least 35% of which are immortalized or cancerous and at least 57% of which are from inflamed tissue. Of particular note is the expression of HURP-8 in libraries made from hematopoietic/immune (30%) and gastrointestinal tissues (22%).

The invention also encompasses HURP variants. A preferred HURP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the HURP amino acid sequence, and which contains at least one functional or structural characteristic of HURP.

The invention also encompasses polynucleotides which encode HURP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence selected from the group consisting of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:16, which encodes a HURP.

The invention also encompasses a variant of a polynucleotide sequence encoding HURP. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HURP. A particular aspect of the invention encompasses a variant a polynucleotide sequence selected from the group SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:16 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:16. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HURP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HURP, some bearing minimal homology to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HURP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode HURP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HURP under

appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HURP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HURP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode HURP and HURP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HURP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, or SEQ ID NO:16 or fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; and Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.)

Methods for DNA sequencing are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp., Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System (GIBCO/BRL, Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ

nucleotide sequence and employing various methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) In particular, genomic DNA is first amplified in the presence of a primer complementary to a linker sequence within the vector and a primer specific to the region predicted to encode the gene. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or another appropriate program to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to 72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.) Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable in that they will include more sequences which contain the 5' regions of genes. Use of a

randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., Genotyper™ and Sequence Navigator™, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HURP may be used in recombinant DNA molecules to direct expression of HURP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express HURP.

As will be understood by those of skill in the art, it may be advantageous to produce HURP-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HURP-encoding sequences for a variety of reasons including, but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random

preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HURP may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of HURP activity, it may be useful to encode a chimeric HURP protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the HURP encoding sequence and the heterologous protein sequence, so that HURP may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding HURP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of HURP, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1983) Proteins, Structures and Molecular Properties, WH Freeman and Co., New York, NY.) Additionally, the amino acid sequence of HURP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active HURP, the nucleotide sequences encoding HURP or derivatives thereof may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HURP and appropriate transcriptional and translational

J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HURP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions, e.g., enhancers, promoters, and 5' and 3' untranslated regions, of the vector and polynucleotide sequences encoding HURP which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters, e.g., hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, La Jolla, CA) or pSport1™ plasmid (GIBCO/BRL), may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding HURP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for HURP. For example, when large quantities of HURP are needed for the induction of antibodies, vectors which direct high level expression of fusion

proteins that are readily purified may be used. Such vectors include, but are not limited to, multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding HURP may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced, and pIN vectors. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) pGEX vectors (Amersham Pharmacia Biotech, Uppsala, Sweden) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. (See, e.g., Ausubel, supra; and Grant et al. (1987) Methods Enzymol. 153:516-544.)

In cases where plant expression vectors are used, the expression of sequences encoding HURP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews. (See, e.g., Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

An insect system may also be used to express HURP. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding HURP may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful

insertion of sequences encoding HURP will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which HURP may be expressed. (See, e.g., Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227.)

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HURP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing HURP in infected host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HURP. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding HURP and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such

modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is preferred. For example, cell lines capable of stably expressing HURP can be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase genes and adenine phosphoribosyltransferase genes, which can be employed in *tk*⁻ or *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *npt* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß glucuronidase and its

substrate GUS, luciferase and its substrate luciferin. Green fluorescent proteins (GFP) (Clontech, Palo Alto, CA) are also used (See, e.g., Chalfie, M. et al. (1994) Science 263:802-805.) These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HURP is inserted within a marker gene sequence, transformed cells containing sequences encoding HURP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HURP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding HURP and express HURP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

The presence of polynucleotide sequences encoding HURP can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding HURP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding HURP to detect transformants containing DNA or RNA encoding HURP.

A variety of protocols for detecting and measuring the expression of HURP, using either polyclonal or monoclonal antibodies specific for the protein, are known in the art.

IV; and Maddox, D.E. et al. (1983) J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HURP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HURP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HURP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HURP may be designed to contain signal sequences which direct secretion of HURP through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding HURP to nucleotide sequences encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA), between the purification domain and the HURP encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing HURP and a nucleic acid encoding 6 histidine

residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on immobilized metal ion affinity chromatography. (IMAC) (See, e.g., Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281.) The enterokinase cleavage site provides a means for purifying HURP from the fusion protein. (See, e.g., Kroll, D.J. et al. (1993) DNA Cell Biol. 12:441-453.)

Fragments of HURP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, T.E. (1984) Protein: Structures and Molecular Properties, pp. 55-60, W.H. Freeman and Co., New York, NY.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of HURP may be synthesized separately and then combined to produce the full length molecule.

**THERAPEUTICS**

Chemical and structural homology exists between HURP-1 and ligatin from mouse (GI 1377880). In addition, HURP-1 is expressed in cancerous, inflamed, fetal/proliferating, and reproductive tissue. Therefore, HURP-1 appears to play a role in cancer, autoimmune/inflammatory disorders, reproductive disorders, and developmental disorders.

Chemical and structural homology exists between HURP-2 and frizzled 6 putative transmembrane receptor from mouse (GI 1151256). In addition, HURP-2 is expressed in cancerous, inflamed, fetal/proliferating, and reproductive tissue. Therefore, HURP-2 appears to play a role in cancer, autoimmune/inflammatory disorders, reproductive disorders, and developmental disorders.

Chemical and structural homology exists among HURP-3, IgE receptor β-subunit protein from rat (GI 204117), and CD20 protein from human (GI 179308). In addition, HURP-3 is expressed in cancerous, inflamed, small intestine, and colon tissue. Therefore,

tissue. Therefore, HURP-4 appears to play a role in cancer, autoimmune/inflammatory disorders, and gastrointestinal disorders.

Chemical and structural homology exists between HURP-5 and CGRP-receptor component protein from mouse (GI 2599066). In addition, HURP-5 is expressed in cancerous and inflamed tissue. Therefore, HURP-5 appears to play a role in cancer, autoimmune/inflammatory disorders, and reproductive disorders.

Chemical and structural homology exists between HURP-6 and signal recognition particle receptor β subunit from mouse (GI 600886). In addition, HURP-6 is expressed in cancerous, inflamed, reproductive, and gastrointestinal tissue. Therefore, HURP-6 appears to play a role in cancer, autoimmune/inflammatory disorders, reproductive disorders, and gastrointestinal disorders.

HURP-7 is expressed in cancerous, inflamed, reproductive and gastrointestinal tissues. Therefore, HURP-7 appears to play a role in cancer, autoimmune/inflammatory disorders, reproductive disorders, and gastrointestinal disorders.

Chemical and structural homology exists between HURP-8 and T-cell receptor β from human (GI 1100182). In addition, HURP-8 is expressed in cancerous, inflamed, hematopoietic/immune, and gastrointestinal tissue. Therefore, HURP-8 appears to play a role in cancer, autoimmune/inflammatory disorders, and gastrointestinal disorders.

Therefore, in one embodiment, an antagonist of HURP may be administered to a subject to treat or prevent a cancer. Such a cancer may include, but is not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds HURP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HURP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HURP may be administered to a subject to treat or prevent a

autoimmune/inflammatory disorder may include, but is not limited to. AIDS, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema. erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, ulcerative colitis, Werner syndrome, and complications of cancer, hemodialysis, and extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma. In one aspect, an antibody which specifically binds HURP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HURP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HURP may be administered to a subject to treat or prevent an autoimmune/inflammatory disorder including, but not limited to, those described above.

In another embodiment, an antagonist of HURP-1, HURP-2, HURP-5, HURP-6, or HURP-7 may be administered to a subject to treat or prevent a reproductive disorder. Such a reproductive disorder may include, but is not limited to, disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, and prostatitis, carcinoma of the male breast and gynecomastia. In one aspect, an antibody

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HURP-1, HURP-2, HURP-5, HURP-6, or HURP-7 may be administered to a subject to treat or prevent a reproductive disorder including, but not limited to, those described above.

In a further embodiment, an antagonist of HURP-3, HURP-4, HURP-6, or HURP-7 may be administered to a subject to treat or prevent a gastrointestinal disorder. Such a gastrointestinal disorder may include, but is not limited to, dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, and AIDS enteropathy. In one aspect, an antibody which specifically binds HURP-3, HURP-4, HURP-6, or HURP-7 may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HURP-3, HURP-4, HURP-6, or HURP-7.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HURP-3, HURP-4, HURP-6, or HURP-7 may be administered to a subject to treat or prevent a gastrointestinal disorder including, but not limited to, those described above.

Therefore, in one embodiment, HURP-8 or a fragment or derivative thereof may be administered to a subject to treat or prevent a gastrointestinal disorder. Such gastrointestinal disorders can include, but are not limited to, dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis,

Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, and AIDS enteropathy.

In another embodiment, a vector capable of expressing HURP-8 or a fragment or derivative thereof may be administered to a subject to treat or prevent a gastrointestinal disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified HURP-8 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a gastrointestinal disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HURP-8 may be administered to a subject to treat or prevent a gastrointestinal disorder including, but not limited to, those listed above.

Therefore, in one embodiment, HURP-1 or HURP-2 or a fragment or derivative thereof may be administered to a subject to treat or prevent a developmental disorder. Such developmental disorders can include, but are not limited to, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome, Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spinal bifida, and congenital glaucoma, cataract, or sensorineural hearing loss.

In another embodiment, a vector capable of expressing HURP-1, HURP-2, or a fragment or derivative thereof may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those described above.

including, but not limited to, those listed above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of HURP may be produced using methods which are generally known in the art. In particular, purified HURP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HURP. Antibodies to HURP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HURP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HURP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small

produced.

Monoclonal antibodies to HURP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HURP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; and Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for HURP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are

well known in the art. Such immunoassays typically involve the measurement of complex formation between HURP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HURP epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding HURP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding HURP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HURP. Thus, complementary molecules or fragments may be used to modulate HURP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HURP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequences complementary to the polynucleotides of the gene encoding HURP. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes encoding HURP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding HURP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding HURP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the

start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HURP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HURP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell

Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HURP, antibodies to HURP, and mimetics, agonists, antagonists, or inhibitors of HURP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may [illegible] pharmaceutically acceptable carriers comprising excipients and auxiliaries

pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be

formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HURP, such labeling would include amount, frequency, and method of administration.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HURP or fragments thereof, antibodies of HURP, and agonists, antagonists or inhibitors of HURP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED50 (the dose therapeutically effective in 50% of the population) or LD50 (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can be expressed as the ED50/LD50 ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 $\mu$g to 100,000 $\mu$g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations,

etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind HURP may be used for the diagnosis of disorders characterized by expression of HURP, or in assays to monitor patients being treated with HURP or agonists, antagonists, or inhibitors of HURP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HURP include methods which utilize the antibody and a label to detect HURP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring HURP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of HURP expression. Normal or standard values for HURP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HURP under conditions suitable for complex formation The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of HURP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HURP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of HURP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HURP, and to monitor regulation of HURP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HURP or closely related molecules may be used to identify nucleic acid sequences which encode HURP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5'

regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding HURP, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the HURP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, or from genomic sequences including promoters, enhancers, and introns of the HURP gene.

Means for producing specific hybridization probes for DNAs encoding HURP include the cloning of polynucleotide sequences encoding HURP or HURP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as $^{32}P$ or $^{35}S$, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HURP may be used for the diagnosis of a disorder associated with expression of HURP. Examples of such a disorder include, but are not limited to, cancers, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and autoimmune/inflammatory disorders such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,

myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, ulcerative colitis, Werner syndrome, and complications of cancer, hemodialysis, and extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma.

Polynucleotide sequences encoding HURP-1, HURP-2, HURP-5, HURP-6, and HURP-7 may be used for the diagnosis of a disorder associated with expression of HURP-1, HURP-2, HURP-5, HURP-6, and HURP-7 respectively. Examples of such a disorder include, but are not limited to, reproductive disorders such as disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, and prostatitis, carcinoma of the male breast and gynecomastia.

Polynucleotide sequences encoding HURP-3, HURP-4, HURP-6, HURP-7, and HURP-8 may be used for the diagnosis of a disorder associated with expression of HURP-3, HURP-4, HURP-6, HURP-7, and HURP-8 respectively. Examples of such a disorder include, but are not limited to, gastrointestinal disorders such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, and AIDS enteropathy.

Polynucleotide sequences encoding HURP-1 and HURP-2 may be used for the

Examples of such a disorder include, but are limited to, developmental disorders such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome, Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spinal bifida, and congenital glaucoma, cataract, or sensorineural hearing loss.

The polynucleotide sequences encoding HURP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered HURP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HURP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding HURP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HURP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HURP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HURP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used.

samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HURP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding HURP, or a fragment of a polynucleotide complementary to the polynucleotide encoding HURP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of HURP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

In one embodiment, the microarray is prepared and used according to methods known in the art. (See, e.g., Chee et al. (1995) PCT application WO95/11995; Lockhart, D. J. et al. (1996) Nat. Biotech. 14:1675-1680; and Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619.)

The microarray is preferably composed of a large number of unique single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs. The oligonucleotides are preferably about 6 to 60 nucleotides in length, more preferably about 15 to 30 nucleotides in length, and most preferably about 20 to 25 nucleotides in length. It may be preferable to use oligonucleotides which are about 7 to 10 nucleotides in length. The microarray may contain oligonucleotides which cover the known 5' or 3' sequence, sequential oligonucleotides which cover the full length sequence, or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray may be oligonucleotides specific to a gene or genes of interest. Oligonucleotides can also be specific to one or more unidentified cDNAs associated with a particular cell type or tissue type. It may be appropriate to use pairs of oligonucleotides on a microarray. The first oligonucleotide in each pair differs from the second oligonucleotide by one nucleotide. This nucleotide is preferably located in the center of the sequence. The second oligonucleotide serves as a control. The number of oligonucleotide pairs may range from about 2 to 1,000,000.

In order to produce oligonucleotides for use on a microarray, the gene of interest is examined using a computer algorithm which starts at the 5' end, or, more preferably, at the 3' end of the nucleotide sequence. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack secondary structure that may interfere with hybridization. In one aspect, the oligomers may be synthesized on a substrate using a light-directed chemical process. (See, e.g., Chee et al., supra.) The substrate may be any suitable solid support, e.g., paper, nylon, any other type of membrane, or a filter, chip, or glass slide.

e.g., Baldeschweiler et al. (1995) PCT application WO95/251116.) An array analogous to a dot or slot blot (HYBRIDOT® apparatus, GIBCO/BRL) may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system or thermal, UV, mechanical, or chemical bonding procedures. An array may also be produced by hand or by using available devices, materials, and machines, e.g. Brinkmann® multichannel pipettors or robotic instruments. The array may contain from 2 to 1,000,000 or any other feasible number of oligonucleotides.

In order to conduct sample analysis using the microarrays, polynucleotides are extracted from a sample. The sample may be obtained from any bodily fluid, e.g., blood, urine, saliva, phlegm, gastric juices, cultured cells, biopsies, or other tissue preparations. To produce probes, the polynucleotides extracted from the sample are used to produce nucleic acid sequences complementary to the nucleic acids on the microarray. If the microarray contains cDNAs, antisense RNAs (aRNAs) are appropriate probes. Therefore, in one aspect, mRNA is reverse-transcribed to cDNA. The cDNA, in the presence of fluorescent label, is used to produce fragment or oligonucleotide aRNA probes. The fluorescently labeled probes are incubated with the microarray so that the probes hybridize to the microarray oligonucleotides. Nucleic acid sequences used as probes can include polynucleotides, fragments, and complementary or antisense sequences produced using restriction enzymes, PCR, or other methods known in the art.

Hybridization conditions can be adjusted so that hybridization occurs with varying degrees of complementarity. A scanner can be used to determine the levels and patterns of fluorescence after removal of any nonhybridized probes. The degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray can be assessed through analysis of the scanned images. A detection system may be used to measure the absence, presence, or level of hybridization for any of the sequences. (See, e.g., Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155.)

In another embodiment of the invention, nucleic acid sequences encoding HURP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial

(See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding HURP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., AT to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, HURP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between HURP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen,

et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with HURP, or fragments thereof, and washed. Bound HURP is then detected by methods well known in the art. Purified HURP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HURP specifically compete with a test compound for binding HURP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HURP.

In additional embodiments, the nucleotide sequences which encode HURP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

## EXAMPLES

For purposes of example, the preparation and sequencing of the MENITUT03 cDNA library, from which Incyte Clone 1255202 was isolated, is described. Preparation and sequencing of cDNAs in libraries in the LIFESEQ™ database have varied over time, and the gradual changes involved use of kits, plasmids, and machinery available at the particular time the library was made and analyzed.

### I. MENITUT03 cDNA Library Construction

The MENITUT03 cDNA library was constructed from malignant meningioma tissue, located in the right cerebellopontine angle of the brain, obtained from a 35-year-old Caucasian female during a right suboccipital craniectomy.

The frozen tissue was homogenized and lysed using a Brinkmann Homogenizer

solution. The lysate was centrifuged over a 5.7 M CsCl cushion using an Beckman SW28 rotor in a Beckman L8-70M Ultracentrifuge (Beckman Instruments) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol pH 4.0, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNAse-free water, and treated with DNase at 37°C. The RNA extraction was repeated twice with acid phenol pH 4.0 and precipitated as before. The mRNA was isolated with the Qiagen Oligotex kit (QIAGEN, Inc.; Chatsworth, CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Cat. #18248-013, Gibco/BRL). The cDNAs were fractionated on a Sepharose CL4B column (Cat. #275105-01, Pharmacia), and those cDNAs exceeding 400 bp were ligated into pINCY 1. The plasmid pINCY 1 was subsequently transformed into DH5α™ competent cells (Cat. #18258-012, Gibco/BRL).

**II. Isolation and Sequencing of cDNA Clones**

Plasmid DNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Catalog #26173; QIAGEN). This kit enabled the simultaneous purification of 96 samples in a 96-well block using multi-channel reagent dispensers. The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, Gibco/BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems; and the reading frame was determined.

**III. Hom logy Searching of cDNA Clones and Their Deduced Proteins**

The nucleotide and amino acid sequences of the Sequence Listing were used to query databases containing previously identified sequences to search for areas of homology (similarity). Several search algorithms and databases were used to help identify identical as well as non-identical sequences.

**HURP-1 and HURP-2**

The first algorithm was originally developed by D.J. Lipman and W.R. Pearson, (1985, Science 227:1435). In this algorithm, the homologous regions of the nucleotide sequences are searched in a two step manner. In the first step, the highest homologous regions are determined by calculating a matching score using a homology score table. The parameter 'Ktup' is used in this step to establish the minimum window size to be shifted for comparing two sequences. Ktup also sets the number of bases that must match to extract the highest homologous region among the sequences. In this step, no insertions or deletions are applied and the homology is displayed as an initial (INIT) value.

In the second step, the homologous regions are aligned to obtain the highest matching score by inserting a gap in order to add a probable deleted portion. The matching score obtained in the first step is recalculated using the homology score Table and the insertion score Table to an optimized (OPT) value in the final output.

DNA homologies between two sequences can be examined graphically using the Harr method of constructing dot matrix homology plots. (Needleman, S.B. and Wunsch, C.O. (1970) J. Mol. Biol 48:443.) This method produces a two-dimensional plot which can be useful in determining regions of homology versus regions of repetition.

The second algorithm was developed by Applied Biosystems Inc. and has been incorporated into the Inherit 670 Sequence Analysis System. In this algorithm, Pattern Specification Language (developed by TRW Inc.) is used to determine regions of homology. There are three parameters that determine how the sequence comparisons are run: window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database is searched for sequences containing regions of homology and the appropriate sequences are scored with an initial value. Subsequently, these homologous regions are examined using dot matrix homology plots to determine regions

Following the search for homologous regions, the sequences from the cDNA clones were classified as to whether they are exact matches (regions of exact homology) homologous human matches (regions of high similarity, but not exact matches), homologous non-human matches (regions of high similarity present in species other than human), or nonmatches (no significant regions of homology to previously identified nucleotide sequences).

Searches of the deduced polypeptides and peptides are done in a manner analogous to that done with the cDNA sequences. The sequence of the polypeptide is used as a query sequence and compared to the previously identified sequences contained in a database such as Swiss/Prot or the NBRF Protein database to find homologous polypeptides. These polypeptides are initially scored for homology using a homology score Table (Orcutt, B.C. and Dayhoff, M.O. Scoring Matrices, PIR Report MAT - 0285 (February 1985)) resulting in an INIT score. The homologous regions are aligned to obtain the highest matching scores by inserting a gap which adds a probable deleted portion. The matching score is recalculated using the homology score Table and the insertion score Table resulting in an optimized (OPT) score. Even in the absence of knowledge of the proper reading frame of an isolated sequence, the above-described polypeptide homology search may be performed by searching all 3 reading frames.

Peptide and protein sequence homologies can also be ascertained using the Inherit 670 Sequence Analysis System in an analogous way to that used in DNA sequence homologies. Pattern Specification Language and parameter windows are used to search polypeptide databases for sequences containing regions of homology which are scored with an initial value. Subsequent examination with a dot-matrix homology plot determines regions of homology versus regions of repetition.

**HURP-3, HURP-4, HURP-5, HURP-6, HURP-7, and HURP-8**

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST (Basic Local Alignment Search Tool). (See e.g. Altschul, S.F. (1993) J. Mol.

BLAST produced alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) Protein Engineering 5:35-51.) The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at $10^{-25}$ for nucleotides and $10^{-8}$ for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and other mammalian sequences (mam), and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for homology.

Additionally, sequences identified from cDNA libraries may be analyzed to identify those gene sequences encoding conserved protein motifs using an appropriate analysis program, e.g., the Block 2 Bioanalysis Program (Incyte, Palo Alto, CA). This motif analysis program, based on sequence information contained in the Swiss-Prot Database and PROSITE, is a method of determining the function of uncharacterized proteins translated from genomic or cDNA sequences. (See, e.g., Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; and Attwood, T. K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.) PROSITE may be used to identify common functional or structural domains in divergent proteins. The method is based on weight matrices. Motifs identified by this method are then calibrated against the SWISS-PROT database in order to obtain a measure of the chance distribution of the matches.

In another alternative, Hidden Markov models (HMMs) may be used to find protein domains, each defined by a dataset of proteins known to have a common biological function. (See, e.g., Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci.

HMMs were initially developed to examine speech recognition patterns, but are now being used in a biological context to analyze protein and nucleic acid sequences as well as to model protein structure. (See, e.g., Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; and Collin, M. et al. (1993) Protein Sci. 2:305-314.) HMMs have a formal probabilistic basis and use position-specific scores for amino acids or nucleotides. The algorithm continues to incorporate information from newly identified sequences to increase its motif analysis capabilities.

## IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; and Ausubel, F.M. et al. supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Homologous molecules are usually identified by selecting those which show

sequences examined in the cDNA library.

## V. Extension of HURP Encoding Polynucleotides

The nucleic acid sequences of Incyte Clones 29167, 150629, 611082, 1223275, 1255202, 1261646, 2083528, and 1451415 were used to design oligonucleotide primers for extending partial nucleotide sequences to full length. For each nucleic acid sequence, one primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences, Plymouth, MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO/BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

| | |
|---|---|
| Step 1 | 94° C for 1 min (initial denaturation) |
| Step 2 | 65° C for 1 min |
| Step 3 | 68° C for 6 min |
| Step 4 | 94° C for 15 sec |
| Step 5 | 65° C for 1 min |
| Step 6 | 68° C for 7 min |
| Step 7 | Repeat steps 4 through 6 for an additional 15 cycles |
| Step 8 | 94° C for 15 sec |
| Step 9 | 65° C for 1 min |
| Step 10 | 68° C for 7:15 min |

A 5 $\mu$l to 10 $\mu$l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQuick™ (QIAGEN Inc.), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 $\mu$l of ligation buffer, 1$\mu$l T4-DNA ligase (15 units) and 1$\mu$l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent E. coli cells (in 40 $\mu$l of appropriate media) were transformed with 3 $\mu$l of ligation mixture and cultured in 80 $\mu$l of SOC medium. (See, e.g., Sambrook, supra, Appendix A, p. 2.) After incubation for one hour at 37° C, the E. coli mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, supra, Appendix A, p. 1) containing 2x carbenicillin (2x carb). The following day, several colonies were randomly picked from each plate and cultured in 150 $\mu$l of liquid LB/2x Carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5 $\mu$l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5 $\mu$l from each sample was transferred into a PCR array.

For PCR amplification, 18 $\mu$l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

| | |
|---|---|
| Step 1 | 94° C for 60 sec |
| Step 2 | 94° C for 20 sec |
| Step 3 | 55° C for 30 sec |
| Step 4 | 72° C for 90 sec |
| Step 5 | Repeat steps 2 through 4 for an additional 29 cycles |
| Step 6 | 72° C for 180 sec |
| Step 7 | 4° C (and holding) |

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:16 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

## VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, and SEQ ID NO:16 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 $\mu$Ci of [$\gamma$-$^{32}$P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified using a Sephadex G-25 superfine resin column (Pharmacia & Upjohn, Kalamazoo, MI). An aliquot containing $10^7$ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

## VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array

appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE™. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; and Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

## VIII. Complementary Polynucleotides

Sequences complementary to the HURP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring HURP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using Oligo 4.06 software and the coding sequence of HURP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HURP-encoding transcript.

## IX. Expression of HURP

Expression of HURP is accomplished by subcloning the cDNA into an appropriate vector and transforming the vector into host cells. This vector contains an appropriate promoter, e.g., ß-galactosidase upstream of the cloning site, operably associated with the cDNA of interest. (See, e.g., Sambrook, supra, pp. 404-433; and

Rosenberg, M. et al. (1983) Methods Enzymol. 101:123-138.)

Induction of an isolated, transformed bacterial strain with isopropyl beta-D-thiogalactopyranoside (IPTG) using standard methods produces a fusion protein which consists of the first 8 residues of ß-galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of HURP into bacterial growth media which can be used directly in the following assay for activity.

## X. Demonstration of HURP Activity

For purposes of example, demonstration of the activity of HURP-8 is described. HURP-8 activity is demonstrated by measuring the binding of HURP-8 to rat brain acetylcholinesterase using affinity chromatography. (Gaston, S.M. et al. (1982) J. Cell. Biochem. 18:447-459.) Acetylcholinesterase/ligatin complexes are purified from rat brain membranes. The complexes are dissociated by dialysis against 5 mM HEPES, 0.5 M EGTA, pH 8.0 buffer overnight at 4°C, and ligatin-free acetylcholinesterase is purified. 200-500 µg HURP-8 is covalently coupled to 1-2 ml of Affi-Gel 10 resin (Bio-Rad Laboratories) at 23°C for two hours in 0.1 M Na phosphate, pH 7.0 buffer. Unreacted Affi-Gel 10 esters are saturated by incubation at 23°C for one hour in 0.1 M glycine, pH 7.0 buffer. The affinity column is equilibrated in 10 mM HEPES, 3 mM $CaCl_2$, 50 mM NaCl, 1 mM $NaN_3$, pH 7.6 buffer. Rat brain acetylcholinesterase is applied to the column. HURP-8 bound acetylcholinesterase can be released from the column by phosphohexose sugars at 10-40 mM.

## XI. Production of HURP Specific Antibodies

HURP substantially purified using PAGE electrophoresis (see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

The HURP amino acid sequence is analyzed using LASERGENE™ software (DNASTAR Inc.) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill

Typically, the oligopeptides are 15 residues in length, and are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel et al. supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

## XII. Purification of Naturally Occurring HURP Using Specific Antibodies

Naturally occurring or recombinant HURP is substantially purified by immunoaffinity chromatography using antibodies specific for HURP. An immunoaffinity column is constructed by covalently coupling anti-HURP antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HURP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HURP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HURP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HURP is collected.

## XIII. Identification of Molecules Which Interact with HURP

HURP, or biologically active fragments thereof, are labeled with $^{125}I$ Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HURP, washed, and any wells with labeled HURP complex are assayed. Data obtained using different concentrations of HURP are used to calculate values for the number, affinity, and association of HURP with the candidate molecules.

Various modifications and variations of the described methods and systems of the

should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8.

2. A substantially purified polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:7.

3. A substantially purified polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:7.

4. A substantially purified polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:8 and a fragment of SEQ ID NO:8.

8. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 7.

9. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 7.

10. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide sequence of claim 7.

11. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, a fragment of SEQ ID NO 9, a fragment of SEQ ID NO:10, a fragment of SEQ ID NO:11, a fragment of SEQ ID NO:12, a fragment of SEQ ID NO:13, a fragment of SEQ ID NO:14, a fragment of SEQ ID NO:15, and a fragment of SEQ ID NO:16.

12. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 11.

13. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 11.

14. An expression vector containing at least a fragment of the polynucleotide of claim 7.

15. A host cell containing the expression vector of claim 14.

16. A method for producing a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID

NO:7, and a fragment of SEQ ID NO:8., the method comprising the steps of:

a) culturing the host cell of claim 15 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

17. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

18. A pharmaceutical composition comprising the polypeptide of claim 4 in conjunction with a suitable pharmaceutical carrier.

19. A pharmaceutical composition comprising the polypeptide of claim 5 in conjunction with a suitable pharmaceutical carrier.

20. A purified antibody which specifically binds to the polypeptide of claim 1.

21. A purified agonist of the polypeptide of claim 1.

22. A purified antagonist of the polypeptide of claim 1.

23. A purified antagonist of the polypeptide of claim 2.

24. A purified antagonist of the polypeptide of claim 3.

27. A method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 23.

28. A method for treating or preventing a gastrointestinal disorder, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 24.

29. A method for treating or preventing a gastrointestinal disorder, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 18.

30. A method for treating or preventing a developmental disorder, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 19.

31. A method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:5, a fragment of SEQ ID NO:6, a fragment of SEQ ID NO:7, and a fragment of SEQ ID NO:8 in a biological sample, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 10 to at least one of the nucleic acids in the biological sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide encoding the polypeptide in the biological sample.

32. The method of claim 31 wherein the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to hybridization.

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
HILLMAN, Jennifer L.
BANDMAN, Olga
YUE, Henry
GUEGLER, Karl J.
CORLEY, Neil C.
AU-YOUNG, Janice
TANG, Y. Tom
SHAH, Purvi
LAL, Preeti
BAUGHN, Mariah

<120> HUMAN RECEPTOR PROTEINS

<130> PF-0480 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/022,939
<151> 1998-02-12

<160> 16

<170> PERL PROGRAM

<210> 1
<211> 584
<212> PRT
<213> Homo sapiens

<220> -
<223> 029167

<400> 1

```
Met Phe Ala Lys Ala Phe Arg Val Lys Ser Asn Thr Ala Ile Lys
  1               5                  10                  15
Gly Ser Asp Arg Arg Lys Leu Arg Ala Asp Val Thr Thr Ala Phe
             20                  25                  30
Pro Thr Leu Gly Thr Asp Gln Val Ser Glu Leu Val Pro Gly Lys
             35                  40                  45
Glu Glu Leu Asn Ile Val Lys Leu Tyr Ala His Lys Gly Asp Ala
             50                  55                  60
Val Thr Val Tyr Val Ser Gly Gly Asn Pro Ile Leu Phe Glu Leu
             65                  70                  75
Glu Lys Asn Leu Tyr Pro Thr Val Tyr Thr Leu Trp Ser Tyr Pro
             80                  85                  90
Asp Leu Leu Pro Thr Phe Thr Thr Trp Pro Leu Val Leu Glu Lys
             95                 100                 105
Leu Val Gly Gly Ala Asp Leu Met Leu Pro Gly Leu Val Met Pro
            110                 115                 120
Pro Ala Gly Leu Pro Gln Val Gln Lys Gly Asp Leu Cys Ala Ile
            125                 130                 135
Ser Leu Val Gly Asn Arg Ala Pro Val Ala Ile Gly Val Ala Ala
            140                 145                 150
```

```
Met Ser Thr Ala Glu Met Leu Thr Ser Gly Leu Lys Gly Arg Gly
                155                 160                 165
Phe Ser Val Leu His Thr Tyr Gln Asp His Leu Trp Arg Ser Gly
                170                 175                 180
Asn Lys Ser Ser Pro Pro Ser Ile Ala Pro Leu Ala Leu Asp Ser
                185                 190                 195
Ala Asp Leu Ser Glu Glu Lys Gly Ser Val Gln Met Asp Ser Thr
                200                 205                 210
Leu Gln Gly Asp Met Arg His Met Thr Leu Glu Gly Glu Glu Glu
                215                 220                 225
Asn Gly Glu Val His Gln Ala Arg Glu Asp Lys Ser Leu Ser Glu
                230                 235                 240
Ala Pro Glu Asp Thr Ser Thr Arg Gly Leu Asn Gln Asp Ser Thr
                245                 250                 255
Asp Ser Lys Thr Leu Gln Glu Gln Met Asp Glu Leu Leu Gln Gln
                260                 265                 270
Cys Phe Leu His Ala Leu Lys Cys Arg Val Lys Lys Ala Asp Leu
                275                 280                 285
Pro Leu Leu Thr Ser Thr Phe Leu Gly Ser His Met Phe Ser Cys
                290                 295                 300
Cys Pro Glu Gly Arg Gln Leu Asp Ile Lys Lys Ser Ser Tyr Lys
                305                 310                 315
Lys Leu Ser Lys Phe Leu Gln Gln Met Gln Gln Glu Gln Ile Ile
                320                 325                 330
Gln Val Lys Glu Leu Ser Lys Gly Val Glu Ser Ile Val Ala Val
                335                 340                 345
Asp Trp Lys His Pro Arg Ile Thr Ser Phe Val Ile Pro Glu Pro
                350                 355                 360
Ser Pro Thr Ser Gln Thr Ile Gln Glu Gly Ser Arg Glu Gln Pro
                365                 370                 375
Tyr His Pro Pro Asp Ile Lys Pro Leu Tyr Cys Val Pro Ala Ser
                380                 385                 390
Met Thr Leu Leu Phe Gln Glu Ser Gly His Lys Lys Gly Ser Phe
                395                 400                 405
Leu Glu Gly Ser Glu Val Arg Thr Ile Val Ile Asn Tyr Ala Lys
                410                 415                 420
Lys Asn Asp Leu Val Asp Ala Asp Asn Lys Asn Leu Val Arg Leu
                425                 430                 435
Asp Pro Ile Leu Cys Asp Cys Ile Leu Glu Lys Asn Glu Gln His
                440                 445                 450
Thr Val Met Lys Leu Pro Trp Asp Ser Leu Leu Thr Arg Cys Leu
                455                 460                 465
Glu Lys Leu Gln Pro Ala Tyr Gln Val Thr Leu Pro Gly Gln Glu
                470                 475                 480
Pro Ile Val Lys Lys Gly Arg Ile Cys Pro Ile Asp Ile Thr Leu
                485                 490                 495
Ala Gln Arg Ala Ser Asn Lys Lys Val Thr Val Val Arg Asn Leu
                500                 505                 510
Glu Ala Tyr Gly Leu Asp Pro Tyr Ser Val Ala Ala Ile Leu Gln
                515                 520                 525
Gln Arg Cys Gln Ala Ser Thr Thr Val Asn Pro Ala Pro Gly Ala
                530                 535                 540
Lys Asp Ser Leu Gln Val Gln Ile Gln Gly Asn Gln Val His His
                545                 550                 555
Leu Gly Trp Leu Leu Leu Glu Glu Tyr Gln Leu Pro Arg Lys His
                560                 565                 570
Ile Gln Gly Leu Glu Lys Ala Leu Lys Pro Gly Lys Lys Lys
                575                 580
```

<210> 2
<211> 706
<212> PRT
<213> Homo sapiens

<220> -
<223> 150629

<400> 2

```
Met Glu Met Phe Thr Phe Leu Leu Thr Cys Ile Phe Leu Pro Leu
  1               5                  10                  15
Leu Arg Gly His Ser Leu Phe Thr Cys Glu Pro Ile Thr Val Pro
             20                  25                  30
Arg Cys Met Lys Met Ala Tyr Asn Met Thr Phe Phe Pro Asn Leu
             35                  40                  45
Met Gly His Tyr Asp Gln Ser Ile Ala Ala Val Glu Met Glu His
             50                  55                  60
Phe Leu Pro Leu Ala Asn Leu Glu Cys Ser Pro Asn Ile Glu Thr
             65                  70                  75
Phe Leu Cys Lys Ala Phe Val Pro Thr Cys Ile Glu Gln Ile His
             80                  85                  90
Val Val Pro Pro Cys Arg Lys Leu Cys Glu Lys Val Tyr Ser Asp
             95                 100                 105
Cys Lys Lys Leu Ile Asp Thr Phe Gly Ile Arg Trp Pro Glu Glu
            110                 115                 120
Leu Glu Cys Asp Arg Leu Gln Tyr Cys Asp Glu Thr Val Pro Val
            125                 130                 135
Thr Phe Asp Pro His Thr Glu Phe Leu Gly Pro Arg Lys Lys Thr
            140                 145                 150
Glu Gln Val Gln Arg Asp Ile Gly Phe Trp Cys Pro Arg His Leu
            155                 160                 165
Lys Thr Ser Gly Gly Gln Gly Tyr Lys Phe Leu Gly Ile Asp Gln
            170                 175                 180
Cys Ala Pro Pro Cys Pro Asn Met Tyr Phe Lys Ser Asp Glu Leu
            185                 190                 195
Glu Phe Ala Lys Ser Phe Ile Gly Thr Val Ser Ile Phe Cys Leu
            200                 205                 210
Cys Ala Thr Leu Phe Thr Phe Leu Thr Phe Leu Ile Asp Val Arg
            215                 220                 225
Arg Phe Arg Tyr Pro Glu Arg Pro Ile Ile Tyr Tyr Ser Val Cys
            230                 235                 240
Tyr Ser Ile Val Ser Leu Met Tyr Phe Ile Gly Phe Leu Leu Gly
            245                 250                 255
Asp Ser Thr Ala Cys Asn Lys Ala Asp Glu Lys Leu Glu Leu Gly
            260                 265                 270
Asp Thr Val Val Leu Gly Ser Gln Asn Lys Ala Cys Thr Val Leu
            275                 280                 285
Phe Met Leu Leu Tyr Phe Phe Thr Met Ala Gly Thr Val Trp Trp
            290                 295                 300
Val Ile Leu Thr Ile Thr Trp Phe Leu Ala Ala Gly Arg Lys Trp
            305                 310                 315
Ser Cys Glu Ala Ile Glu Gln Lys Ala Val Trp Phe His Ala Val
            320                 325                 330
Ala Trp Gly Thr Pro Gly Phe Leu Thr Val Met Leu Leu Ala Met
            335                 340                 345
```

```
Asn Lys Val Glu Gly Asp Asn Ile Ser Gly Val Cys Phe Val Gly
                350                 355                 360
Leu Tyr Asp Leu Asp Ala Ser Arg Tyr Phe Val Leu Leu Pro Leu
                365                 370                 375
Cys Leu Cys Val Phe Val Gly Leu Ser Leu Leu Leu Ala Gly Ile
                380                 385                 390
Ile Ser Leu Asn His Val Arg Gln Val Ile Gln His Asp Gly Arg
                395                 400                 405
Asn Gln Glu Lys Leu Lys Lys Phe Met Ile Arg Ile Gly Val Phe
                410                 415                 420
Ser Gly Leu Tyr Leu Val Pro Leu Val Thr Leu Leu Gly Cys Tyr
                425                 430                 435
Val Tyr Glu Gln Val Asn Arg Ile Thr Trp Glu Ile Thr Trp Val
                440                 445                 450
Ser Asp His Cys Arg Gln Tyr His Ile Pro Cys Pro Tyr Gln Ala
                455                 460                 465
Lys Ala Lys Ala Arg Pro Glu Leu Ala Leu Phe Met Ile Lys Tyr
                470                 475                 480
Leu Met Thr Leu Ile Val Gly Ile Ser Ala Val Phe Trp Val Gly
                485                 490                 495
Ser Lys Lys Thr Cys Thr Glu Trp Ala Gly Phe Phe Lys Arg Asn
                500                 505                 510
Arg Lys Arg Asp Pro Ile Ser Glu Ser Arg Arg Val Leu Gln Glu
                515                 520                 525
Ser Cys Glu Phe Phe Leu Lys His Asn Ser Lys Val Lys His Lys
                530                 535                 540
Lys Lys His Tyr Lys Pro Ser Ser His Lys Leu Lys Val Ile Ser
                545                 550                 555
Lys Ser Met Gly Thr Ser Thr Gly Ala Thr Ala Asn His Gly Thr
                560                 565                 570
Ser Ala Val Ala Ile Thr Ser His Asp Tyr Leu Gly Gln Glu Thr
                575                 580                 585
Leu Thr Glu Ile Gln Thr Ser Pro Glu Thr Ser Met Arg Glu Val
                590                 595                 600
Lys Ala Asp Gly Ala Ser Thr Pro Arg Leu Arg Glu Gln Asp Cys
                605                 610                 615
Gly Glu Pro Ala Ser Pro Ala Ala Ser Ile Ser Arg Leu Ser Gly
                620                 625                 630
Glu Gln Val Asp Gly Lys Gly Gln Ala Gly Ser Val Ser Glu Ser
                635                 640                 645
Ala Arg Ser Glu Gly Arg Ile Ser Pro Lys Ser Asp Ile Thr Asp
                650                 655                 660
Thr Gly Leu Ala Gln Ser Asn Asn Leu Gln Val Pro Ser Ser Ser
                665                 670                 675
Glu Pro Ser Ser Leu Lys Gly Ser Thr Ser Leu Leu Val His Pro
                680                 685                 690
Val Ser Gly Val Arg Lys Glu Gln Gly Gly Gly Cys His Ser Asp
                695                 700                 705
Thr
```

<210> 3
<211> 267
<212> PRT
<213> Homo sapiens

<220> -
<223> 611082

<400> 3
Met Met Ser Ser Lys Pro Thr Ser His Ala Glu Val Asn Glu Thr
1 5 10 15
Ile Pro Asn Pro Tyr Pro Pro Ser Ser Phe Met Ala Pro Gly Phe
20 25 30
Gln Gln Pro Leu Gly Ser Ile Asn Leu Glu Asn Gln Ala Gln Gly
35 40 45
Ala Gln Arg Ala Gln Pro Tyr Gly Ile Thr Ser Pro Gly Ile Phe
50 55 60
Ala Ser Ser Gln Pro Gly Gln Gly Asn Ile Gln Met Ile Asn Pro
65 70 75
Ser Val Gly Thr Ala Val Met Asn Phe Lys Glu Glu Ala Lys Ala
80 85 90
Leu Gly Val Ile Gln Ile Met Val Gly Leu Met His Ile Gly Phe
95 100 105
Gly Ile Val Leu Cys Leu Ile Ser Phe Ser Phe Arg Glu Val Leu
110 115 120
Gly Phe Ala Ser Thr Ala Val Ile Gly Gly Tyr Pro Phe Trp Gly
125 130 135
Gly Leu Ser Phe Ile Ile Ser Gly Ser Leu Ser Val Ser Ala Ser
140 145 150
Lys Glu Leu Ser Arg Cys Leu Val Lys Gly Ser Leu Gly Met Asn
155 160 165
Ile Val Ser Ser Ile Leu Ala Phe Ile Gly Val Ile Leu Leu Leu
170 175 180
Val Asp Met Cys Ile Asn Gly Val Ala Gly Gln Asp Tyr Trp Ala
185 190 195
Val Leu Ser Gly Lys Gly Ile Ser Ala Thr Leu Met Ile Phe Ser
200 205 210
Leu Leu Glu Phe Phe Val Ala Cys Ala Thr Ala His Phe Ala Asn
215 220 225
Gln Ala Asn Thr Thr Thr Asn Met Ser Val Leu Val Ile Pro Asn
230 235 240
Met Tyr Glu Ser Asn Pro Val Thr Pro Ala Ser Ser Ser Ala Pro
245 250 255
Pro Arg Cys Asn Asn Tyr Ser Ala Asn Ala Pro Lys
260 265

<210> 4
<211> 248
<212> PRT
<213> Homo sapiens

<220> -
<223> 1223275

<400> 4
Met Thr Ser Gln Pro Val Pro Asn Glu Thr Ile Ile Val Leu Pro
1 5 10 15
Ser Asn Val Ile Asn Phe Ser Gln Ala Glu Lys Pro Glu Pro Thr
20 25 30
Asn Gln Gly Gln Asp Ser Leu Lys Lys His Leu His Ala Glu Ile

```
                35                  40                  45
Lys Val Ile Gly Thr Ile Gln Ile Leu Cys Gly Met Met Val Leu
                50                  55                  60
Ser Leu Gly Ile Ile Leu Ala Ser Ala Ser Phe Ser Pro Asn Phe
                65                  70                  75
Thr Gln Val Thr Ser Thr Leu Leu Asn Ser Ala Tyr Pro Phe Ile
                80                  85                  90
Gly Pro Phe Phe Phe Ile Ile Ser Gly Ser Leu Ser Ile Ala Thr
                95                  100                 105
Glu Lys Arg Leu Thr Lys Leu Leu Val His Ser Ser Leu Val Gly
                110                 115                 120
Ser Ile Leu Ser Ala Leu Ser Ala Leu Val Gly Phe Ile Ile Leu
                125                 130                 135
Ser Val Lys Gln Ala Thr Leu Asn Pro Ala Ser Leu Gln Cys Glu
                140                 145                 150
Leu Asp Lys Asn Asn Ile Pro Thr Arg Ser Tyr Val Ser Tyr Phe
                155                 160                 165
Tyr His Asp Ser Leu Tyr Thr Thr Asp Cys Tyr Thr Ala Lys Ala
                170                 175                 180
Ser Leu Ala Gly Thr Leu Ser Leu Met Leu Ile Cys Thr Leu Leu
                185                 190                 195
Glu Phe Cys Leu Ala Val Leu Thr Ala Val Leu Arg Trp Lys Gln
                200                 205                 210
Ala Tyr Ser Asp Phe Pro Gly Ser Val Leu Phe Leu Pro His Ser
                215                 220                 225
Tyr Ile Gly Asn Ser Gly Met Ser Ser Lys Met Thr His Asp Cys
                230                 235                 240
Gly Tyr Glu Glu Leu Leu Thr Ser
                245
```

<210> 5
<211> 130
<212> PRT
<213> Homo sapiens

<220>
<221> unsure
<222> 120
<223> unknown, or other

<220> -
<223> 1255202

<400> 5

```
Met Glu Val Lys Asp Ala Asn Ser Ala Leu Leu Ser Asn Tyr Glu
  1               5                  10                  15
Val Phe Gln Leu Leu Thr Asp Leu Lys Glu Gln Arg Lys Glu Ser
                20                  25                  30
Gly Lys Asn Lys His Ser Ser Gly Gln Gln Asn Leu Asn Thr Ile
                35                  40                  45
Thr Tyr Glu Thr Leu Lys Tyr Ile Ser Lys Thr Pro Cys Arg His
                50                  55                  60
Gln Ser Pro Glu Ile Val Arg Glu Phe Leu Thr Ala Leu Lys Ser
                65                  70                  75
His Lys Leu Thr Lys Ala Glu Lys Leu Gln Leu Leu Asn His Arg
```

```
                80                  85                  90
Pro Val Thr Ala Val Glu Ile Gln Leu Met Val Glu Glu Ser Glu
                95                 100                 105
Glu Arg Leu Thr Glu Glu Gln Ile Glu Ala Leu Leu His Thr Xaa
               110                 115                 120
Asn Gln His Ser Ala Cys Arg Thr Arg Arg
               125                 130
```

<210> 6
<211> 271
<212> PRT
<213> Homo sapiens

<220> -
<223> 1261646

<400> 6

```
Met Ala Ser Ala Asp Ser Arg Arg Val Ala Asp Gly Gly Gly Ala
  1               5                  10                  15
Gly Gly Thr Phe Gln Pro Tyr Leu Asp Thr Leu Arg Gln Glu Leu
                20                  25                  30
Gln Gln Thr Asp Pro Thr Leu Leu Ser Val Val Val Ala Val Leu
                35                  40                  45
Ala Val Leu Leu Thr Leu Val Phe Trp Lys Leu Ile Arg Ser Arg
                50                  55                  60
Arg Ser Ser Gln Arg Ala Val Leu Leu Val Gly Leu Cys Asp Ser
                65                  70                  75
Gly Lys Thr Leu Leu Phe Val Arg Leu Leu Thr Gly Leu Tyr Arg
                80                  85                  90
Asp Thr Gln Thr Ser Ile Thr Asp Ser Cys Ala Val Tyr Arg Val
                95                 100                 105
Asn Asn Asn Arg Gly Asn Ser Leu Thr Leu Ile Asp Leu Pro Gly
               110                 115                 120
His Glu Ser Leu Arg Leu Gln Phe Leu Glu Arg Phe Lys Ser Ser
               125                 130                 135
Ala Arg Ala Ile Val Phe Val Val Asp Ser Ala Ala Phe Gln Arg
               140                 145                 150
Glu Val Lys Asp Val Ala Glu Phe Leu Tyr Gln Val Leu Ile Asp
               155                 160                 165
Ser Met Gly Leu Lys Asn Thr Pro Ser Phe Leu Ile Ala Cys Asn
               170                 175                 180
Lys Gln Asp Ile Ala Met Ala Lys Ser Ala Lys Leu Ile Gln Gln
               185                 190                 195
Gln Leu Glu Lys Glu Leu Asn Thr Leu Arg Val Thr Arg Ser Ala
               200                 205                 210
Ala Pro Ser Thr Leu Asp Ser Ser Ser Thr Ala Pro Ala Gln Leu
               215                 220                 225
Gly Lys Lys Gly Lys Glu Phe Glu Phe Ser Gln Leu Pro Leu Lys
               230                 235                 240
Val Glu Phe Leu Glu Cys Ser Ala Lys Gly Gly Arg Gly Asp Val
```

<210> 7
<211> 591
<212> PRT
<213> Homo sapiens

<220> -
<223> 2083528

<400> 7

```
Met Leu Leu Ala Thr Leu Leu Leu Leu Leu Leu Gly Gly Ala Leu
  1               5                  10                  15
Ala His Pro Asp Arg Ile Ile Phe Pro Asn His Ala Cys Glu Asp
             20                  25                  30
Pro Pro Ala Val Leu Leu Glu Val Gln Gly Thr Leu Gln Arg Pro
             35                  40                  45
Leu Val Arg Asp Ser Arg Thr Ser Pro Ala Asn Cys Thr Trp Leu
             50                  55                  60
Ile Leu Gly Ser Lys Glu Gln Thr Val Thr Ile Lys Phe Gln Lys
             65                  70                  75
Leu His Leu Ala Cys Gly Ser Glu Arg Leu Thr Leu Arg Ser Pro
             80                  85                  90
Leu Gln Pro Leu Ile Ser Leu Cys Glu Ala Pro Pro Ser Pro Leu
             95                 100                 105
Gln Leu Pro Gly Gly Asn Val Thr Ile Thr Tyr Ser Tyr Ala Gly
            110                 115                 120
Ala Arg Ala Pro Met Gly Gln Gly Phe Leu Leu Ser Tyr Ser Gln
            125                 130                 135
Asp Trp Leu Met Cys Leu Gln Glu Glu Phe Gln Cys Leu Asn His
            140                 145                 150
Arg Cys Val Ser Ala Val Gln Arg Cys Asp Gly Val Asp Ala Cys
            155                 160                 165
Gly Asp Gly Ser Asp Glu Ala Gly Cys Ser Ser Asp Pro Phe Pro
            170                 175                 180
Gly Leu Thr Pro Arg Pro Val Pro Ser Leu Pro Cys Asn Val Thr
            185                 190                 195
Leu Glu Asp Phe Tyr Gly Val Phe Ser Ser Pro Gly Tyr Thr His
            200                 205                 210
Leu Ala Ser Val Ser His Pro Gln Ser Cys His Trp Leu Leu Asp
            215                 220                 225
Pro His Asp Gly Arg Arg Leu Ala Val Arg Phe Thr Ala Leu Asp
            230                 235                 240
Leu Gly Phe Gly Asp Ala Val His Val Tyr Asp Gly Pro Gly Pro
            245                 250                 255
Pro Glu Ser Ser Arg Leu Leu Arg Ser Leu Thr His Phe Ser Asn
            260                 265                 270
Gly Lys Ala Val Thr Val Glu Thr Leu Ser Gly Gln Ala Val Val
            275                 280                 285
Ser Tyr His Thr Val Ala Trp Ser Asn Gly Arg Gly Phe Asn Ala
            290                 295                 300
Thr Tyr His Val Arg Gly Tyr Cys Leu Pro Trp Asp Arg Pro Cys
            305                 310                 315
Gly Leu Gly Ser Gly Leu Gly Ala Gly Glu Gly Leu Gly Glu Arg
            320                 325                 330
Cys Tyr Ser Glu Ala Gln Arg Cys Asp Gly Ser Trp Asp Cys Ala
            335                 340                 345
Asp Gly Thr Asp Glu Glu Asp Cys Pro Gly Cys Pro Pro Gly His
```

```
                350                 355                 360
Phe Pro Cys Gly Ala Ala Gly Thr Ser Gly Ala Thr Ala Cys Tyr
                365                 370                 375
Leu Pro Ala Asp Arg Cys Asn Tyr Gln Thr Phe Cys Ala Asp Gly
                380                 385                 390
Ala Asp Glu Arg Arg Cys Arg His Cys Gln Pro Gly Asn Phe Arg
                395                 400                 405
Cys Arg Asp Glu Lys Cys Val Tyr Glu Thr Trp Val Cys Asp Gly
                410                 415                 420
Gln Pro Asp Cys Ala Asp Gly Ser Asp Glu Trp Asp Cys Ser Tyr
                425                 430                 435
Val Leu Pro Arg Lys Val Ile Thr Ala Ala Val Ile Gly Ser Leu
                440                 445                 450
Val Cys Gly Leu Leu Leu Val Ile Ala Leu Gly Cys Thr Cys Lys
                455                 460                 465
Leu Tyr Ala Ile Arg Thr Gln Glu Tyr Ser Ile Phe Ala Pro Leu
                470                 475                 480
Ser Arg Met Glu Ala Glu Ile Val Gln Gln Gln Ala Pro Pro Ser
                485                 490                 495
Tyr Gly Gln Leu Ile Ala Gln Gly Ala Ile Pro Pro Val Glu Asp
                500                 505                 510
Phe Pro Thr Glu Asn Pro Asn Asp Asn Ser Val Leu Gly Asn Leu
                515                 520                 525
Arg Ser Leu Leu Gln Ile Leu Arg Gln Asp Met Thr Pro Gly Gly
                530                 535                 540
Gly Pro Gly Ala Arg Arg Arg Gln Arg Gly Arg Leu Met Arg Arg
                545                 550                 555
Leu Val Pro Val Ser Ala Ala Gly Ala Cys Ser Leu Glu Pro Thr
                560                 565                 570
Pro Arg Leu Gly Pro Leu Arg Pro Asp Pro Arg Ser His Leu Leu
                575                 580                 585
Leu Leu Pro Leu Arg Pro
                590
```

<210> 8
<211> 314
<212> PRT
<213> Homo sapiens

<220> -
<223> 1451415

<400> 8

```
Met Gly Thr Arg Leu Leu Cys Trp Val Val Leu Gly Phe Leu Gly
  1                 5                  10                  15
Thr Asp His Thr Gly Ala Gly Val Ser Gln Ser Pro Arg Tyr Lys
                 20                  25                  30
Val Ala Lys Arg Gly Gln Asp Val Ala Leu Arg Cys Asp Pro Ile
                 35                  40                  45
Ser Gly His Val Ser Leu Phe Trp Tyr Gln Gln Ala Leu Gly Gln
                 50                  55                  60
Gly Pro Glu Phe Leu Thr Tyr Phe Gln Asn Glu Ala Gln Leu Asp
                 65                  70                  75
Lys Ser Gly Leu Pro Ser Asp Arg Phe Phe Ala Glu Arg Pro Glu
                 80                  85                  90
```

```
Gly Ser Val Ser Thr Leu Lys Ile Gln Arg Thr Gln Gln Glu Asp
 95                  100                 105
Ser Ala Val Tyr Leu Cys Ala Ser Ser Leu Ser Pro Ala Gly Val
110                 115                 120
Val Asp Thr Gln Tyr Phe Gly Pro Gly Thr Arg Leu Thr Val Leu
125                 130                 135
Glu Asp Leu Lys Asn Val Phe Pro Pro Glu Val Ala Val Phe Glu
140                 145                 150
Pro Ser Glu Ala Glu Ile Ser His Thr Gln Lys Ala Thr Leu Val
155                 160                 165
Cys Leu Ala Thr Gly Phe Tyr Pro Asp His Val Glu Leu Ser Trp
170                 175                 180
Trp Val Asn Gly Lys Glu Val His Ser Gly Val Ser Thr Asp Pro
185                 190                 195
Gln Pro Leu Lys Glu Gln Pro Ala Leu Asn Asp Ser Arg Tyr Cys
200                 205                 210
Leu Ser Ser Arg Leu Arg Val Ser Ala Thr Phe Trp Gln Asn Pro
215                 220                 225
Arg Asn His Phe Arg Cys Gln Val Gln Phe Tyr Gly Leu Ser Glu
230                 235                 240
Asn Asp Glu Trp Thr Gln Asp Arg Ala Lys Pro Val Thr Gln Ile
245                 250                 255
Val Ser Ala Glu Ala Trp Gly Arg Ala Asp Cys Gly Phe Thr Ser
260                 265                 270
Glu Ser Tyr Gln Gln Gly Val Leu Ser Ala Thr Ile Leu Tyr Glu
275                 280                 285
Ile Leu Leu Gly Lys Ala Thr Leu Tyr Ala Val Leu Val Ser Ala
290                 295                 300
Leu Val Leu Met Ala Met Val Lys Arg Lys Asp Ser Arg Gly
305                 310
```

<210> 9
<211> 2005
<212> DNA
<213> Homo sapiens

<220> -
<223> 029167

<400> 9

```
cgccccagca tggctgcccc cacggctgag ggcctggcag ctgctgcgcc ctcgctttct  60
tgacattccc tggcttctgt gctctcttcc ccaggccacc ccagcagaca tgtttgccaa 120
ggcctttcgg gtcaagtcca acacggccat caaggggtcg gacaggagaa agcttcgagc 180
tgatgtgaca actgctttcc ccacccttgg aactgatcaa gtctctgagt tagtacctgg 240
aaaggaggag ctcaacattg tgaagttgta tgctcacaaa ggggatgcag tgactgtgta 300
cgtgagtggt ggtaacccca tcctctttga actggagaaa aatctgtatc caacagtgta 360
cacgctgtgg tcctatcctg atcttctgcc aacctttaca acatggcctc tggtgctcga 420
gaaactggta gggggagcag atttgatgct gcctggactg gtgatgcccc ctgctggtct 480
gcctcaggta cagaagggcg acctctgtgc catttctttg gtggggaaca gagcccctgt 540
agccattgga gttgcagcca tgtccacagc tgagatgctc acgtctggcc tgaagggaag 600
```

ggatgagctg ttacagcaat gcttcttaca tgccttgaag tgccgagtca aaaaggctga 960
cctcccttta ctcaccagca ctttccttgg cagccacatg ttctcctgct gccccgaagg 1020
acgacaactg gacataaaga agtcaagcta caaaaagctc tctaagttcc tgcagcaaat 1080
gcagcaggag cagattatac aggtgaagga gctgagcaaa ggggtggaga gcattgtggc 1140
tgtggactgg aaacacccga ggattacatc tttcgtcata cccgagccct ccccgacctc 1200
ccagactatc caggagggta gcagggaaca gccctatcac cctccagata taaaacccct 1260
ctactgtgtc ccagccagca tgaccctgct cttccaggag tctggccaca agaaggggag 1320
ctttctggag ggcagtgagg tccgaacgat cgtcattaac tacgccaaga aaaatgacct 1380
ggttgatgca gacaacaaaa atcttgtgag attggatccc atcctatgtg actgcatctt 1440
agagaaaaat gaacagcata cagtcatgaa gcttccatgg gacagtcttc tgaccaggtg 1500
tttggaaaaa ttacagcctg cctatcaagt gacccttccc ggacaagagc ccattgtgaa 1560
gaaagggaga atctgtccaa ttgacatcac cctagcacaa agagcgtcta ataaaaaggt 1620
gaccgtggtc cggaacttgg aggcctatgg tctggaccca tactcagtgg ctgccatcct 1680
tcagcagcga tgccaggcta gcaccaccgt caatcctgcc cctggggcca aggacagcct 1740
tcaggtgcag atccagggaa accaggtcca ccacctcggc tggctattgc ttgaagagta 1800
tcagctccct cgaaaacaca tccaaggtct agaaaaggcc ctcaaacctg gcaagaagaa 1860
gtgacagact cttttgtctc acgtggtgga tccggtggaa atccaagctc tgggctggta 1920
atttttatga gcattttcag cttttgcaaa tacaaaatat aattctttac aaaaataaat 1980
ttttattcta atctaaaaaa aaaaa 2005

<210> 10
<211> 2833
<212> DNA
<213> Homo sapiens

<220> -
<223> 150629

<400> 10
ccttcgaccg ccccccgagt aattgaccca ggactcattt tcaggaaagc ctgaaaatga 60
gtaaaatagt gaaatgagga atttgaacat tttatctttg gatggggatc ttctgaggat 120
gcaaagagtg attcatccaa gccatgtggt aaaatcagga atttgaagaa aatggagatg 180
tttacatttt tgttgacgtg tatttttcta cccctcctaa gagggcacag tctcttcacc 240
tgtgaaccaa ttactgttcc cagatgtatg aaaatggcct acaacatgac gtttttccct 300
aatctgatgg gtcattatga ccagagtatt gccgcggtgg aaatggagca ttttcttcct 360
ctcgcaaatc tggaatgttc accaaacatt gaaactttcc tctgcaaagc atttgtacca 420
acctgcatag aacaaattca tgtggttcca ccttgtcgta aactttgtga gaaagtatat 480
tctgattgca aaaaattaat tgacactttt gggatccgat ggcctgagga gcttgaatgt 540
gacagattac aatactgtga tgagactgtt cctgtaactt ttgatccaca cacagaattt 600
cttggtcctc ggaagaaaac agaacaagtc caaagagaca ttggattttg gtgtccaagg 660
catcttaaga cttctggggg acaaggatat aagtttctgg gaattgacca gtgtgcgcct 720
ccatgcccca acatgtattt taaaagtgat gagctagagt ttgcaaaaag ttttattgga 780
acagtttcaa tattttgtct ttgtgcaact ctgttcacat tccttacttt tttaattgat 840
gttagaagat tcagataccc agagagacca attatatatt actctgtctg ttacagcatt 900
gtatctctta tgtacttcat tggatttttg ctgggcgata gcacagcctg caataaggca 960
gatgagaagc tagaacttgg tgacactgtt gtcctaggct ctcaaaataa ggcttgcacc 1020
gttttgttca tgcttttgta ttttttcaca atggctggca ctgtgtggtg ggtgattctt 1080
accattactt ggttcttagc tgcaggaaga aaatggagtt gtgaagccat cgagcaaaaa 1140
gcagtgtggt ttcatgctgt tgcatgggga acaccaggtt tcctgactgt tatgcttctt 1200
gctatgaaca aagttgaagg agacaacatt agtggagttt gctttgttgg cctttatgac 1260
ctggatgctt ctcgctactt tgtactcttg ccactgtgcc tttgtgtgtt tgttgggctc 1320
tctcttcttt tagctggcat tatttcctta aatcatgttc gacaagtcat acaacatgat 1380

gctgggtttt ttaaacgaaa tcgcaagaga gatccaatca gtgaaagtcg aagagtacta 1740
caggaatcat gtgagttttt cttaaagcac aattctaaag ttaaacacaa aaagaagcac 1800
tataaaccaa gttcgcacaa gctgaaggtc atttccaaat ccatgggaac cagcacagga 1860
gctacagcaa atcatggcac ttctgcagta gcaattacta gccatgatta cctaggacaa 1920
gaaactttga cagaaatcca aacctcacca gaaacatcaa tgagagaggt gaaagcggac 1980
ggagctagca cccccaggtt aagagaacag gactgtggtg aacctgcctc gccagcagca 2040
tccatctcca gactctctgg ggaacaggtc gacgggaagg gccaggcagg cagtgtatct 2100
gaaagtgcgc ggagtgaagg aaggattagt ccaaagagtg atattactga cactggcctg 2160
gcacagagca acaatttgca ggtccccagt tcttcagaac caagcagcct caaaggttcc 2220
acatctctgc ttgttcaccc agtttcagga gtgagaaaag agcagggagg tggttgtcat 2280
tcagatactt gaagaacatt ttctctcgtt actcagaagc aaatttgtgt tacactggaa 2340
gtgacctatg cactgttttg taagaatcac tgttacgttc ttcttttgca cttaaagttg 2400
cattgcctac tgttatactg gaaaaaatag agttcaagaa taatatgact catttcacac 2460
aaaggttaat gacaacaata tacctgaaaa cagaaatgtg caggttaata atatttttt 2520
aatagtgtgg gaggacagag ttagaggaat cttccttttc tatttatgaa gattctactc 2580
ttggtaagag tattttaaga tgtactatgc tattttactt ttttgatata aaatcaagat 2640
atttctttgc tgaagtattt aaatcttatc cttgtatctg ttatacatat gtgaaataag 2700
cttataagga attgagctgt ttggaatgct attcgagtat ttgaccatgc gaatgtgaga 2760
tttagcattt ggtagcttta acggaatttc aagaaatgga aaagagctct ttatccgtaa 2820
gaaggttgcc aag 2833

<210> 11
<211> 1317
<212> DNA
<213> Homo sapiens

<220> -
<223> 611082

<400> 11
aaggaacaaa gtaagtccat tgatacgttc ttgcctatct ctcctccaaa tcaatgggca 60
caaactgtgg ctggtctacc tgtgtgggtt ctgttctcta gattggaggg atgaagacaa 120
gttcttgact ctatgttgag gccagttgaa aaatgaggga gaataaaacc atgaacgaaa 180
caagaaagaa acaaaacaga agaggaatga aaaagacata atgatgtcat ccaagccaac 240
aagccatgct gaagtaaatg aaaccatacc caacccttac ccaccaagca gctttatggc 300
tcctggattt caacagcctc tgggttcaat caacttagaa aaccaagctc agggtgctca 360
gcgtgctcag ccctacggca tcacatctcc gggaatcttt gctagcagtc aaccgggtca 420
aggaaatata caaatgataa atccaagtgt gggaacagca gtaatgaact ttaaagaaga 480
agcaaaggca ctaggggtga tccagatcat ggttggattg atgcacattg gttttggaat 540
tgttttgtgt ttaatatcct tctcttttag agaagtatta ggttttgcct ctactgctgt 600
tattggtgga tacccattct ggggtggcct ttcttttatt atctctggct ctctctctgt 660
gtcagcatcc aaggagcttt cccgttgtct ggtgaaaggc agcctgggaa tgaacattgt 720
tagttctatc ttggccttca ttggagtgat tctgctgctg gtggatatgt gcatcaatgg 780
ggtagctggc caagactact gggccgtgct ttctggaaaa ggcatttcag ccacgctgat 840
gatcttctcc ctcttggagt tcttcgtagc ttgtgccaca gcccattttg ccaaccaagc 900
aaacaccaca accaatatgt ctgtcctggt tattccaaat atgtatgaaa gcaaccctgt 960
gacaccagcg tcttcttcag ctcctcccag atgcaacaac tactcagcta atgcccctaa 1020
atagtaaaag aaaaaggggt atcagtctaa tctcatggag aaaaactact tgcaaaaact 1080
tcttaagaag atgtctttta ttgtctacaa tgatttctag tctttaaaaa ctgtgtttga 1140
gatttgtttt taggttggtc gctaatgatg gctgtatctc ccttcactgt ctcttcctac 1200
attaccacta ctacatgctg gcaaaggtga aggatcagag gactgaaaaa tgattctgca 1260
actctcttaa agttagaaat gtttctgttc atattacttt ttccttaata aaatgtc 1317

<210> 12
<211> 1522
<212> DNA

<213> Homo sapiens

<220> -
<223> 1223275

<400> 12
acaaatctac aatctgtgag tatcacatcc tgtatagctg taaacactgg aataaggaag 60
ggctgatgac tttcagaaga tgaaggtaag tagaaaccgt tgatgggact gagaaaccag 120
agttaaaacc tctttggagc ttctgaggac tcagctggaa ccaacgggca cagttggcaa 180
caccatcatg acatcacaac ctgttcccaa tgagaccatc atagtgctcc catcaaatgt 240
catcaacttc tcccaagcag agaaacccga acccaccaac caggggcagg atagcctgaa 300
gaaacatcta cacgcagaaa tcaaagttat tgggactatc cagatcttgt gtggcatgat 360
ggtattgagc ttggggatca ttttggcatc tgcttccttc tctccaaatt ttacccaagt 420
gacttctaca ctgttgaact ctgcttaccc attcatagga ccctttttt ttatcatctc 480
tggctctcta tcaatcgcca cagagaaaag gttgaccaag cttttggtgc atagcagcct 540
ggttggaagc attctgagtg ctctgtctgc cctggtgggt ttcattatcc tgtctgtcaa 600
acaggccacc ttaaatcctg cctcactgca gtgtgagttg gacaaaaata atataccaac 660
aagaagttat gtttcttact tttatcatga ttcactttat accacggact gctatacagc 720
caaagccagt ctggctggaa ctctctctct gatgctgatt tgcactctgc tggaattctg 780
cctagctgtg ctcactgctg tgctgcggtg gaaacaggct tactctgact tccctgggag 840
tgtacttttc ctgcctcaca gttacattgg taattctggc atgtcctcaa aaatgactca 900
tgactgtgga tatgaagaac tattgacttc ttaagaaaaa agggagaaat attaatcaga 960
aagttgattc ttatgataat atggaaaagt taaccattat agaaaagcaa agcttgagtt 1020
tcctaaatgt aagcttttaa agtaatgaac attaaaaaaa accattattt cactgtcatt 1080
taagatatgt gttcattggg gatctcttga tttgcctgac attgacttca gcaaaagcac 1140
ggggctgtaa attaccattt actagattag ccaaatagtc tgaatttcca gaaaacaagg 1200
cagaatgatc attcccagaa acatttccca gaaaatgttt cccagaaaac tagacagaat 1260
gatcattcaa tggatcacag tgaagcaaag gacacaactt tttattgtac cccttaattg 1320
tcaacaggag ttaactgatt tgttgtggtg ctcagacttt tttatacagg tgctagtgtt 1380
ttatcctatg tattttaact cattagtgca taaaggcaag ccccatataa tgaagtctca 1440
gggtatatga aagtagctgg cttcaaaata aaatttttga gtgcaaaaaa aaaaaaaaaa 1500
aaaaaaaaag gggcggacgg cc 1522

<210> 13
<211> 542
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 455, 502, 539
<223> a or g or c or t, unknown, or other

<220> -
<223> 1255202

<400> 13
ggactgggca ccttggcgct gttggtggcg gcggagacag ctgtgaagtg tgaggttctt 60
gtctgctggc agctaggggc gacgaggcgg gacgtcatgg aagtgaagga tgccaattct 120
gcgcttctca gtaactacga ggtatttcag ttactaactg atctgaaaga gcagcgtaaa 180
gaaagtggaa agaataaaca cagctctggg caacagaact tgaacactat cacctatgaa 240
acgttaaaat acatatcaaa aacaccatgc aggcaccaga gtcctgaaat tgtcagagaa 300
tttctcacag cattgaaaag ccacaagttg accaaagctg agaagctcca gctgctgaac 360
caccggcctg tgactgctgt ggagatccag ctgatggtgg aagagagtga agagcggctc 420
acggaggagc agattgaagc tcttctccac accgncaacc agcattctgc ctgcagaacc 480
agaaggtgag ccggaggagg antacaaaca gccattgtgg gagtctctca acgtcacana 540
tt 542

<210> 14
<211> 1870
<212> DNA
<213> Homo sapiens

<220> -
<223> 1261646

<400> 14
gactacggag actatggctt aggaaccacc attcgcgtga ggctctgctc cccgcgcctg 60
cgcagagtgc agggccacgt cgcttttgct gtaccgggga ccacgcgtct catccatggc 120
ttccgcggac tcgcgccggg tggcagatgg cggcggtgcc gggggcacct tccagcccta 180
cctagacacc ttgcggcagg agctgcagca gacggaccca acgctgttgt cagtagtggt 240
ggcggttctt gcggtgctgc tgacgctagt cttctggaag ttaatccgga gcagaaggag 300
cagtcagaga gctgttcttc ttgttggcct ttgtgattcc gggaaaacgt tgctctttgt 360
caggttgtta acaggccttt atagagacac tcagacgtcc attactgaca gctgtgctgt 420
atacagagtc aacaataaca ggggcaatag tctgaccttg attgaccttc ccggccatga 480
gagtttgagg cttcagttct tagagcggtt taagtcttca gccagggcta ttgtgtttgt 540
tgtggatagt gcagcattcc agcgagaggt gaaagatgtg gctgagtttc tgtatcaagt 600
cctcattgac agtatgggtc tgaagaatac accatcattc ttaatagcct gcaataagca 660
agatattgca atggcaaaat cagcaaagtt aattcaacag cagctggaga aagaactcaa 720
caccttacga gttacccgtt ctgctgcccc cagcacactg gacagttcca gcactgcccc 780
tgctcagctg gggaagaaag gcaaagagtt tgaattctca cagttgcccc tcaaagtgga 840
gttcctggag tgcagtgcca agggtggaag aggggacgtg ggctctgctg acatccagga 900
cttggagaaa tggctggcta aaattgcctg agaggcagct ctaaagcaca agacctggat 960
gtgtgacaca cagttttgga aaaaggtctg tggtagtctg gagttgatga ggaaggggta 1020
caagatgtgg ttagaaacat ttctttgttc tggaaacaaa gtactgttga aaccagcttg 1080
gaatttttt ttttttttt tttaagttca gttctccctt atggctgcct ttcaaacaag 1140
taccttttat ctgatgcctg tatcttccct ttgttaaggt gtaacttgat gtagggtcaa 1200
ggtttttgtg acaacaggca gactccacac agagaggata tgatgagaat atggccatca 1260
cctgaaaagt tttcttatct tctgtgcttt tggtccctgg aaacaaatcc gcctatgtat 1320
gaagctagtt gatttccagt tgcactattt ccagttgcct ctgaagttca caggcaatac 1380
attgtctagt cctttgcgaa tttctctgat ttgtgggcac agttatgaag tttccccaca 1440
tgtgaagaca ggtacaaaat agcagagcca agcagacagt gggtctattc ttcattagct 1500
cagtgacttg tccacactcg tcttagcact tacgtttcaa aagcttgtca caaacccttg 1560
gagtcattcc cagataatag aactggaaat gataaatccc ctaatgccaa gggtctagtg 1620
tgttcttagt ggttatactg ggaagtgtgt ggagatttag gtgctgctct gctgctctgg 1680
atggctgaag gctcctgggc catcttcatg tgctgcttga agagctccta ttttgtactc 1740
ctggctagaa tgctgtggaa caaatacaaa gtgaaaaaag ttctctgtag atttctgaag 1800
tgcatattca ttgatgccaa gaaaaaaaaa aagttgcctt tttgaaaaaa aaaaaaaaaa 1860
aacgctcgag 1870

<210> 15
<211> 3014
<212> DNA
<213> Homo sapiens

<220> -
<223> 2083528

<400> 15
caccgccccc actcccgggc tgccgccgcc tccccgcccc cagccctggc atccagagta 60
cgggtcgagc ccgggccatg gagcccccct ggggaggcgg caccagggag cctgggcgcc 120
cggggctccg ccgcgacccc atcgggtaga ccacagaagc tccgggaccc ttccggcacc 180
tctggacagc acaggatgct gttggccacc ctcctcctcc tcctccttgg aggcgctctg 240
gcccatccag accggattat ttttccaaat catgcttgtg aggacccccc agcagtgctc 300

ttagaagtgc agggcacctt acagaggccc ctggtccggg acagccgcac ctcccctgcc 360
aactgcacct ggctcatcct gggcagcaag gaacagactg tcaccatcaa gttccagaag 420
ctacacctgg cctgtggctc agagcgctta accctacgct cccctctcca gccactgatc 480
tccctgtgtg aggcacctcc cagcccctctg cagctgcccg ggggcaacgt caccatcact 540
tacagctatg ctggggccag agcacccatg ggccagggct tcctgctctc ctacagccaa 600
gattggctga tgtgcctgca ggaagagttt cagtgcctga accaccgctg tgtatctgct 660
gtccagcgct gtgatggggt tgatgcctgt ggcgatggct ctgatgaagc aggttgcagc 720
tcagaccccc tccctggcct gaccccaaga cccgtccccc ccctgccttg caatgtcacc 780
ttggaggact tctatggggt cttctcctct cctggatata cacacctagc ctcagtctcc 840
caccccccagt cctgccattg gctgctggac ccccatgatg gccggcggct ggccgtgcgc 900
ttcacagctc tggacttggg ctttggagat gcagtgcatg tgtatgacgg ccctgggccc 960
cctgagagct cccgactact gcgtagtctc acccacttca gcaatggcaa ggctgtcact 1020
gtggagacac tgtctggcca ggctgttgtg tcctaccaca cagttgcttg gagcaatggt 1080
cgtggcttca atgccaccta ccatgtgcgg ggctattgct tgccttggga cagaccctgt 1140
ggcttaggct ctggcctggg agctggcgaa ggcctaggtg agcgctgcta cagtgaggca 1200
cagcgctgtg acggctcatg ggactgtgct gacggcacag atgaggagga ctgcccaggc 1260
tgcccacctg gacacttccc ctgtggggct gctggcacct ctggtgccac agcctgctac 1320
ctgcctgctg accgctgcaa ctaccagact ttctgtgctg atggagcaga tgagagacgc 1380
tgtcggcatt gccagcctgg caatttccga tgccgggacg agaagtgcgt gtatgagacg 1440
tgggtgtgcg atgggcagcc agactgtgcg gacggcagtg atgagtggga ctgctcctat 1500
gttctgcccc gcaaggtcat tacagctgca gtcattggca gcctagtgtg cggcctgctc 1560
ctggtcatcg ccctgggctg cacctgcaag ctctatgcca ttcgcaccca ggagtacagc 1620
atctttgccc ccctctcccg gatggaggct gagattgtgc agcagcaggc acccccttcc 1680
tacgggcagc tcattgccca gggtgccatc ccacctgtag aagactttcc tacagagaat 1740
cctaatgata actcagtgct gggcaacctg cgttctctgc tacagatctt acgccaggat 1800
atgactccag gaggtggccc aggtgcccgc cgtcgtcagc ggggccgctt gatgcgacgc 1860
ctggtacccg tctccgccgc tggggcttgc tccctcgaac caacaccccg gctcgggcct 1920
ctgaggccag atcccaggtc acaccttctg ctgctccccct tgaggcccta gatggtggca 1980
caggtccagc ccgtgagggc ggggcagtgg gtgggcaaga tggggagcag gcacccccac 2040
tgcccatcaa ggctcccctc ccatctgcta gcacgtctcc agcccccact actgtccctg 2100
aagccccagg gccactgccc tcactgcccc tagagccatc actattgtct ggagtggtgc 2160

<220> -
<223> 1451415

<400> 16
agacaaggac ggcccttatc ctctgcttct gttcacagtg acactgatct ggtaaagccc 60
ccatcctggc ctgaccctgc catgggcacc aggctcctct gctgggtggt cctgggtttc 120
ctagggacag atcacacagg tgctggagtc tcccagtccc ctaggtacaa agtcgcaaag 180
agaggacagg atgtagctct caggtgtgat ccaatttcgg gtcatgtatc ccttttttgg 240
taccaacagg ccctggggca ggggccagag tttctgactt atttccagaa tgaagctcaa 300
ctagacaaat cggggctgcc cagtgatcgc ttctttgcag aaaggcctga gggatccgtc 360
tccactctga agatccagcg cacacagcag gaggactccg ccgtgtatct ctgtgccagc 420
agcttaagcc cagcgggagt tgtagatacg cagtattttg gcccaggcac ccggctgaca 480
gtgctcgagg acctgaaaaa cgtgttccca cccgaggtcg ctgtgtttga gccatcagaa 540
gcagagatct cccacaccca aaaggccaca ctggtgtgcc tggccacagg cttctacccc 600
gaccacgtgg agctgagctg gtgggtgaat gggaaggagg tgcacagtgg ggtcagcaca 660
gacccgcagc ccctcaagga gcagcccgcc ctcaatgact ccagatactg cctgagcagc 720
cgcctgaggg tctcggccac cttctggcag aaccccccgca accacttccg ctgtcaagtc 780
cagttctacg ggctctcgga gaatgacgag tggacccagg atagggccaa acctgtcacc 840
cagatcgtca gcgccgaggc ctggggtaga gcagactgtg gcttcacctc cgagtcttac 900
cagcaagggg tcctgtctgc caccatcctc tatgagatct tgctagggaa ggccaccttg 960
tatgccgtgc tggtcagtgc cctcgtgctg atggccatgg tcaagagaaa ggattccaga 1020
ggctagctcc aaaaccatcc caggtcattc ttcatcctca cccaggattc tcctgtacct 1080
gctcccaatc tgtgttccta aaagtgattc tcactctgct tctcatctcc tacttacatg 1140
aatacttctc tctttttttct gtttccctga agattgagct cccaaccccn aagtacgaaa 1200
taggctaaac caataaaaaa ttgtgtgttg ggcctggttg catt 1244